LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

numero 177
maggio 1983
anno XVI
volume XXX

# La Legge del mare

*Nel dicembre scorso 119 nazioni hanno firmato la Convenzione sulla Legge del mare. Gli Stati Uniti non hanno aderito, ma i programmi di sviluppo della nuova legge procederanno ugualmente*

di Elisabeth Mann Borgese

Il 6 dicembre dello scorso anno i rappresentanti di 119 nazioni, riuniti a Montego Bay in Giamaica, hanno firmato la Convenzione delle Nazioni Unite sulla Legge del mare. La cerimonia ha coronato oltre 10 anni di lavori della più numerosa e più lunga di tutte le conferenze internazionali. La terza Conferenza delle Nazioni Unite sulla Legge del mare aveva impegnato ben 3000 delegati in 15 sessioni per un totale di 585 giorni. Iniziata nel caos e nello scontro di interessi politici ed economici opposti, la Conferenza è riuscita, una bozza dopo l'altra, a redigere una «costituzione per gli oceani» in 320 articoli (organizzati in 17 parti principali) e nove allegati tecnici.

Per quanto sia possibile sintetizzare in una sola frase il succo di questa enorme mole di lavoro, si può dire che la Convenzione sostituisce al tradizionale sistema del *laissez faire* per quanto concerne la libertà dei mari un sistema di gestione completamente nuovo. In luogo dello sconfinato mare bidimensionale, la Convenzione si occupa di una risorsa tridimensionale circoscritta, le cui profondità hanno lo stesso interesse economico della superficie. Essa pone il 40 per cento del mare e i fondali adiacenti alle coste dei continenti e delle isole sotto l'amministrazione degli stati titolari delle coste in questione. Riserva l'altro 60 per cento della superficie e le acque sottostanti alla tradizionale libertà dei mari, ma trasferisce le ricchezze dei fondali marini, pari al 42 per cento della superficie terrestre, al «patrimonio comune del genere umano». La Convenzione assoggetta tale patrimonio alla gestione di un ente internazionale per i fondali marini, completamente nuovo, la International Seabed Authority, che ha la capacità di creare reddito, il potere di imporre tasse e una specie di predominio sulla tecnologia per lo sfruttamento dei mari.

Proprio in opposizione alla Seabed Authority, gli Stati Uniti, e altre 15 nazioni industrializzate hanno ritirato la loro adesione alla Convenzione. Queste nazioni peraltro hanno firmato il documento finale *(final act)* della Conferenza. (La Turchia, invece, non ha aderito né all'una né all'altro.) Le nazioni che hanno firmato il documento finale, ma non la Convenzione, hanno la facoltà di intervenire, in veste di osservatori non partecipanti, alle future riunioni degli enti creati dalla Convenzione. La firma della Convenzione dà a questa forza di legge, a patto che venga ratificata da almeno 60 nazioni. Una volta che la Convenzione avrà ottenuto tale ratifica, è prevedibile che anche i non firmatari si decideranno a firmare come partecipanti a pieno titolo.

La Convenzione prevede un ordinamento globale organico per la tutela dell'ambiente marino, un nuovo regime per le ricerche scientifiche sul mare e un sistema esauriente per risolvere le vertenze. Essa assicura libertà di navigazione e libero transito negli stretti che servono alla navigazione internazionale, appellandosi al nuovo concetto legale di «passaggio di transito», un diritto che non può essere sospeso in nessuna circostanza. Aggiorna e codifica il diritto tradizionale. Definisce, infine, i diritti e le responsabilità degli stati e delle organizzazioni internazionali nel promuovere e intensificare un nuovo ordine nei mari e negli oceani.

La costituzione della nuova Legge del mare tocca tutti i problemi che la comunità mondiale deve affrontare nella nostra epoca: cibo ed energia, corsa agli armamenti, comunicazioni e «rivoluzione nell'informazione», scambi commerciali, politiche dei prodotti del suolo, gestione e conservazione delle risorse, integrazione regionale e quegli orientamenti della scienza e della tecnologia che sono alla base di molte di tali questioni. È come se tutti i problemi fluissero insieme nello spazio oceanico e quest'ultimo fosse diventato il nostro laboratorio per la costruzione di un nuovo ordine mondiale. Un ordine che - si spera - si dimostri più razionale, più umano e più sensibile alle effettive necessità del mondo di quanto lo sia stato il vecchio ordine che si va disgregando nella fame e nella violenza.

Fino al 1973, quando ebbe inizio la terza Conferenza delle Nazioni Unite sulla Legge del mare, questa era sempre stata fatta da uno o due o, al massimo, una mezza dozzina di stati marinari. Nel corso dei secoli, o dei millenni, i mari del Vecchio Mondo furono sottoposti prima al codice di navigazione di Hammurabi e poi al diritto marittimo rodiota. Il Mediterraneo, culla della civiltà occidentale, diventò il *mare nostrum* di Roma. Nel 1493, la Spagna e il Portogallo si divisero fra loro lo spazio oceanico che si estendeva oltre lo Stretto di Gibilterra. Per la potenza marittima dell'Inghilterra del Seicento, John

**L'adozione della Convenzione sulla Legge del mare è stata ripresa in questa fotografia del quartier generale delle Nazioni Unite a New York il 30 aprile dello scorso anno. Sopra i delegati c'è un tabellone elettronico che riporta il risultato della votazione: 130 nazioni favorevoli, 4 contrarie e 17 astenute. Il 6 dicembre scorso, 119 di queste nazioni hanno firmato a Montego Bay, in Giamaica, la Convenzione, e 143 hanno firmato il documento finale della Conferenza sulla Legge del mare. Sottoscrivendo quest'ultimo, ma non la Convenzione, una nazione ha il diritto di intervenire come osservatore non votante alle future riunioni della Commissione preparatoria.**

Selden propose la teoria del *mare clausum,* il mare chiuso «di proprietà» degli stati rivieraschi che, come ai loro tempi la Spagna e il Portogallo, avrebbero avuto il diritto di controllarlo e di esigere un pedaggio dai navigatori di altre nazioni. Contro questa dottrina, il giurista olandese Hugo Grotius propose il concetto di *mare liberum,* il mare libero aperto ai navigatori di tutte le nazioni, ognuna delle quali avrebbe concesso lo stesso diritto alle altre. Nella realtà storica la libertà dei mari poté essere goduta soltanto da chi era in grado di farla rispettare: in tutta la storia moderna solo un piccolo numero di paesi dell'Europa occidentale e dell'America Settentrionale.

Due drammatici sviluppi nelle vicende umane, concomitanti e spesso in conflitto fra loro, resero indispensabile la convocazione della terza Conferenza delle Nazioni Unite sulla Legge del mare. Il primo di essi è costituito dalla dimensione marittima della rivoluzione scientifico-industriale che sta faticosamente trasformando le società industrializzate e che sta incidendo profondamente sui rapporti fra queste società e quelle preindustrializzate. La scoperta del fenomeno dell'espansione del fondo marino ha avuto la conseguenza letteralmente sismica di porre come idea ordinatrice centrale in geologia la tettonica a zolle. Questa stessa scoperta ha implicazioni economiche di vasta portata. Là dove, lungo gli assi di espansione, il fondo marino si sta spaccando, i metalli si sono separati selettivamente dal mantello terrestre fino alla superficie. Si sa oggi che questo processo è responsabile della deposizione dei filoni di metalli non ferrosi fin qui estratti nelle miniere in roccia dura sparse nel mondo. Dalla stessa più approfondita conoscenza del fondo marino si sa che non solo nel fondo bentonico, ma anche nelle piattaforme continentali, esso contiene più petrolio e più gas dei trappi esistenti nella roccia continentale (gran parte della quale, naturalmente, era anch'essa un tempo fondo marino).

Nel frattempo, il rapido sviluppo tecnologico dell'industria della pesca ha annullato il vecchio concetto secondo il quale le risorse ittiche dell'oceano sono inesauribili. Nel medesimo tempo, la presunta infinità dell'oceano è stata sottoposta a dura prova dall'inquinamento dovuto all'urbanizzazione, all'industrializzazione e alla crescita costante della popolazione.

Così le premesse della vecchia Legge del mare si erosero e la legge incominciò a sgretolarsi. Gli Stati Uniti diedero l'avvio all'aumento delle rivendicazioni nazionali sul mare e sulle sue risorse nel 1945 con le *Proclamations* di Truman sulla piattaforma continentale e la zona di pesca allargata. Subito si accese una mischia per suddividere lo spazio oceanico, proprio come un secolo prima era stata suddivisa l'Africa.

Mentre la Rivoluzione Industriale penetrava nei mari dall'emisfero settentrionale, nell'emisfero meridionale andava acquistando forza il secondo grande impulso che avrebbe poi dato forma alla nuova Legge del mare. Dagli imperi coloniali creati dalle ex potenze marittime europee nacque tutta una serie di nuove nazioni sovrane. Alla prima Conferenza delle Nazioni Unite sulla Legge del mare, indetta nel 1958, o alla seconda, indetta nel 1960, avevano preso parte non più di 70 nazioni. Nel 1973, quando si riunì la terza Conferenza, il numero dei membri delle Nazioni Unite era più che raddoppiato. All'Assemblea generale la maggioranza era passata alle nazioni povere dell'Asia, dell'Africa e dell'America Latina, le quali erano decise a far sentire la propria voce nelle deliberazioni della Conferenza. Di lì a poco avrebbe preso consistenza la loro richiesta di un nuovo ordine economico internazionale. La nuova Legge del mare non solo avrebbe fatto parte di questo nuovo ordine, ma gli avrebbe fatto perfino da modello.

La Convenzione sulla Legge del mare dimostra chiaramente di essere stata plasmata dalle rivoluzioni parallele, ma spesso intersecantisi, che si sono avute nella tecnologia e nelle relazioni internazionali. È stata la rivoluzione tecnologica a stimolare dal Nord l'espansione della giurisdizione nazionale nello spazio oceanico, mentre è stata la rivoluzione avvenuta nelle relazioni internazionali a premere, dal Sud, perché si cercassero nuove, forti istituzioni internazionali. Le due rivoluzioni sono sempre in primo piano nei voluminosi verbali della lunga Conferenza. Peraltro la Convenzione stessa e i verbali della Conferenza rispecchiano anche la presenza di numerose personalità, sia del Nord sia del Sud, abili, erudite e di grande statura politica: per citarne solo alcuni, Arvid Pardo della Repubblica di Malta; Hamilton Shirley Amerasinghe dello Sri Lanka, presidente della Conferenza fino alla sua prematura scomparsa, avvenuta nel 1981; Jens Evensen della Norvegia; Jorge Castañeda del Messico; S. P. Jagota dell'India e Louis Sohn degli Stati Uniti. Il loro talento individuale ha interagito con le pressioni collettive esercitate sia dal Nord sia dal Sud per pilotare i lavori della Conferenza.

Fu Pardo, ad esempio, a proporre in un memorabile discorso pronunciato nel 1967 davanti al primo Comitato delle Nazioni Unite, che le ricchezze dei fondali marini fossero dichiarate patrimonio comune del genere umano. Né la libertà dei mari né la sovranità che a quel tempo veniva rivendicata sulle acque marine dagli stati costieri, egli disse, avrebbero potuto scacciare il triplice spettro dell'inquinamento, dell'esaurimento della vita marina e della discordia internazionale. Soltanto la cooperazione internazionale avrebbe potuto salvaguardare la pace e le ricchezze dei mari, e la cosa sarebbe stata favorita dal fatto che, in essi, gli interessi costituiti non erano ancora saldamente radicati come lo erano sulla terraferma.

Fu nominato immediatamente un Comitato sugli usi pacifici dei fondali marini; nel 1970, fu approvata una Dichiarazione di princìpi, che includeva la tesi principale di Pardo e, nel 1972, la limitazione dei fondali marini a usi pacifici fu inclusa nel trattato che metteva al bando le armi nucleari dal fondo del mare. Nel 1972, il Comitato per i fondali marini, tuttora all'opera, chiese la convocazione della terza Conferenza delle Nazioni Unite sulla Legge del mare e incominciò a preparare l'ordine dei lavori.

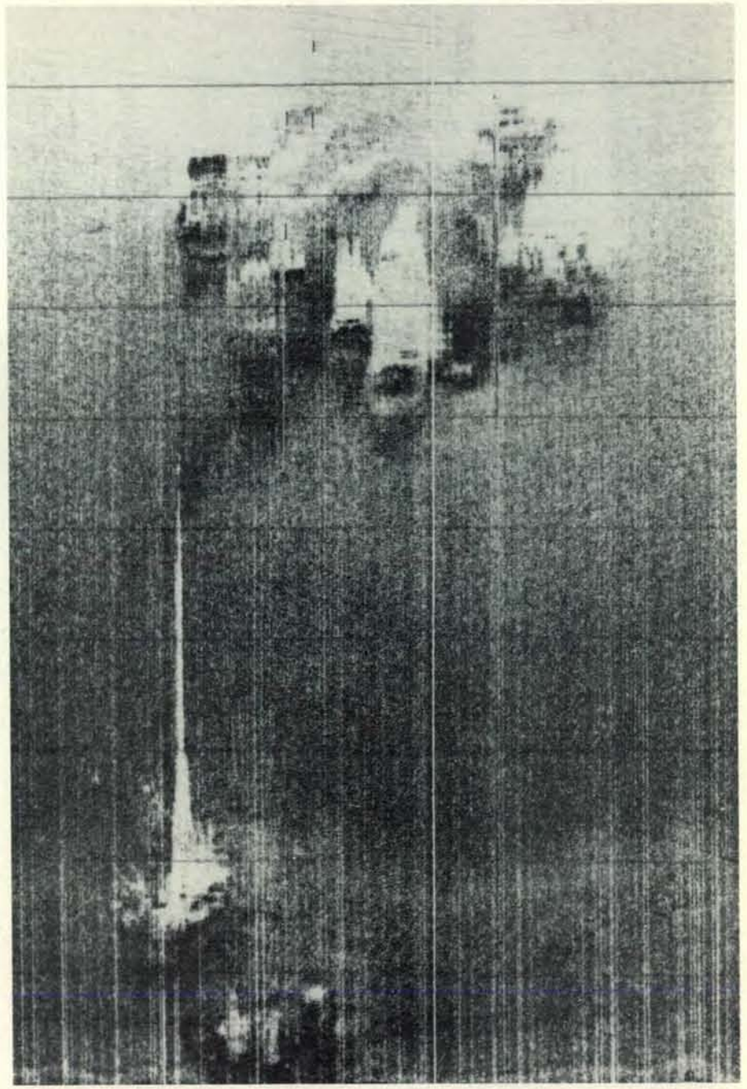
Questo fu completato nel 1973, allorché ebbero inizio i lavori della Conferenza sulla Legge del mare. Organizzata in un Comitato generale, in tre principali comitati di lavoro e in una rete intricata di gruppi regionali, gruppi d'interesse, gruppi incaricati di condurre i negoziati, riunioni plenarie formali e informali e un comitato addetto alla stesura delle bozze in sei lingue, questa conferenza è stata la più grande e la più complessa della storia.

Il compito di delimitare gli spazi oceanici - le acque territoriali tradizionali e la zona contigua nonché le nuove zone economiche allargate e le piattaforme continentali - fu assegnato al potente secondo Comitato. Qui si dovettero conciliare gli interessi contrastanti di tre gruppi di nazioni. Gli stati rivieraschi, naturalmente, cercavano di estendere il più possibile i limiti della loro giurisdizione e di assicurarsi il massimo dei diritti ottenibili nella zona sotto tale giurisdizione. All'altra estremità dello spettro c'erano gli stati senza sbocchi sul mare, fra cui numerosi paesi africani, i più poveri fra i poveri. Essi non potevano rivendicare nessuna zona e rischiavano di perdere, nei mari che si sarebbero chiusi per loro, anche quei pochi diritti di cui forse godevano, soprattutto riguardo alla pesca. In mezzo stavano gli stati marinari, che condividevano alcuni degli interessi degli stati rivieraschi ma, dipendendo pure dalla libertà dei mari per poter impiegare le proprie flotte in acque lontane, condividevano anche gli interessi degli stati senza sbocchi sul mare. Ognuno di questi gruppi era una complessa aggregazione di paesi sviluppati e in via di sviluppo sia a economia di mercato sia a pianificazione centrale.

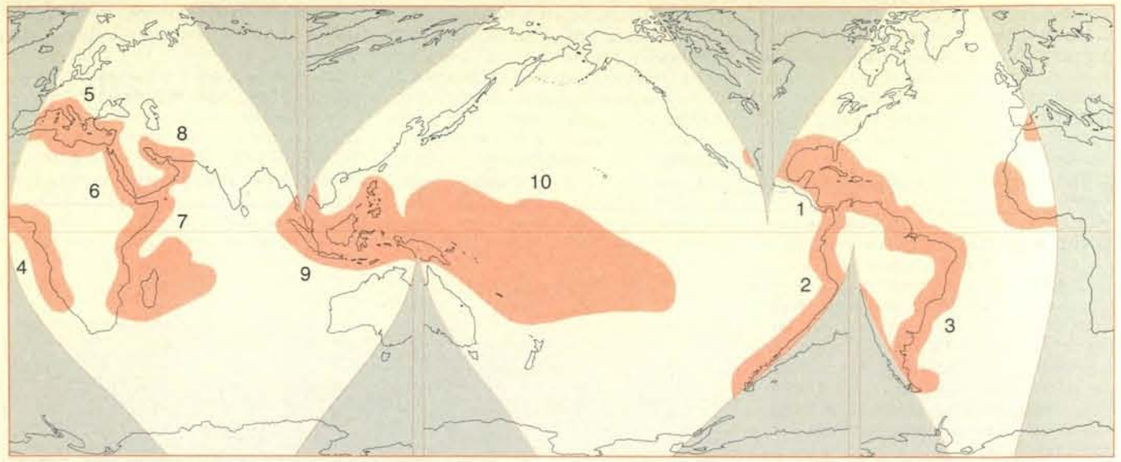
Non può esserci il benché minimo dubbio che a trarre vantaggio dall'espansione della giurisdizione nazionale siano stati di gran lunga i paesi industrializzati, soprattutto gli Stati Uniti, il Canada, l'Australia e il Sudafrica, oltre ad alcuni grandi paesi in via di sviluppo, fra cui l'India, il Brasile e il Messico. Gli articoli della Convenzione, promulgati dal secondo Comitato, confermano il limite di 12 miglia per le acque territoriali oltre a una zona contigua di altre 12 miglia. Al di là di questi limiti essi accordano agli stati rivieraschi diritti sovrani sull'uso delle risorse naturali nella zona economica di 200 miglia, recentemente sancita, e inoltre molti altri diritti in merito a determinate attività, come per esempio la ricerca scientifica. Dove la piattaforma continentale si estende al di là della zona economica di 200 miglia, con essa si estendono, però non oltre le 350 miglia, anche i diritti relativi alle risorse della piattaforma.

Lo stato rivierasco non ha peraltro sovranità territoriale nella propria zona economica, così come l'ha sulla terraferma e sulle acque territoriali limitrofe. Questo nuovo tipo di sovranità funzionale potrebbe essere considerato la forma meno maligna di ingrandimento nazionale. Si tratta di un concetto nuovo e dinamico del diritto internazionale, che troverà indubbiamente nuove applicazioni in futuro.

La creazione della zona economica



**Depositi di minerali nelle profondità marine del Golfo di California sono visibili in questa immagine ripresa con un sistema sonar a visione laterale. Le caratteristiche geomorfologiche che emergono dal fondo sono monticelli e «pinnacoli», tipici delle zone di sorgenti idrotermali, che si trovano in prossimità dei centri di espansione del fondo oceanico. In queste aree, dall'acqua calda emessa dal fondo precipitano solfuri ricchi di metalli. Le caratteristiche geomorfologiche si elevano per 30 metri al di sopra del fondo. La distanza fra quattro linee orizzontali è di 450 metri. La fotografia è stata fornita da Fred N. Spiess e Peter F. Lonsdale della Scripps Institution of Oceanography.**




**Vari «piani d'azione» sono già in vigore o in preparazione nell'ambito dell'Environment Programme delle Nazioni Unite. I piani in vigore si riferiscono al Mediterraneo *(5)*, al Mar Rosso e al Golfo di Aden *(6)* e alla Kuwait Action Plan Region *(8)*; quelli in preparazione, ai Caraibi *(1)*, al Pacifico sudorientale *(2)*, all'Atlantico sud-occidentale *(3)*, all'Africa occidentale *(4)*, all'Africa orientale *(7)*, ai mari dell'Asia orientale *(9)* e al Pacifico sudoccidentale *(10)*. La proiezione cartografica è la stessa della cartina alle pagine 18 e 19.**

risponde anche alla necessità di gestione dell'ambiente marittimo. Essa costituisce un incentivo di carattere economico all'azione delle singole nazioni a questo riguardo, ma non soddisfa assolutamente la necessità in questione, perché né l'inquinamento né i pesci rispettano i confini nazionali.

Dal lato negativo, il concetto di zona economica è stato definito il più grande arraffare territoriale della storia. L'espansione della giurisdizione nazionale nei mari accresce certamente la disuguaglianza fra le nazioni. Se si sperava che la Conferenza ponesse limiti chiari e inequivocabili alla giurisdizione nazionale, bisogna dire che essa è venuta meno anche da questo punto di vista. Così come vengono definiti, i limiti sono elastici, con scappatoie che permetteranno di avanzare ulteriori rivendicazioni in qualsiasi momento. È facile prevedere che la scoperta di qualsiasi risorsa di notevole entità in un punto qualsiasi dello spazio oceanico porterà a una rivendicazione da parte dello stato insulare o rivierasco più vicino, nonostante le solenni dichiarazioni di non appropriabilità da parte di qualsiasi stato o persona (come potrebbe essere, per esempio, una grande società per azioni) del patrimonio comune del genere umano «oltre i limiti della giurisdizione nazionale».

Nondimeno, si può sperare che futuri sviluppi portino a ovviare a tutti i punti deboli e a mitigare le ingiustizie relative a questa parte della Convenzione. Con tutta probabilità ciò avverrà attraverso un progresso nella cooperazione regionale e il rafforzamento delle organizzazioni internazionali. Come si vedrà, c'è ragione di attendersi una certa forma di spartizione dei benefici economici, derivanti dallo sfruttamento della zona economica.

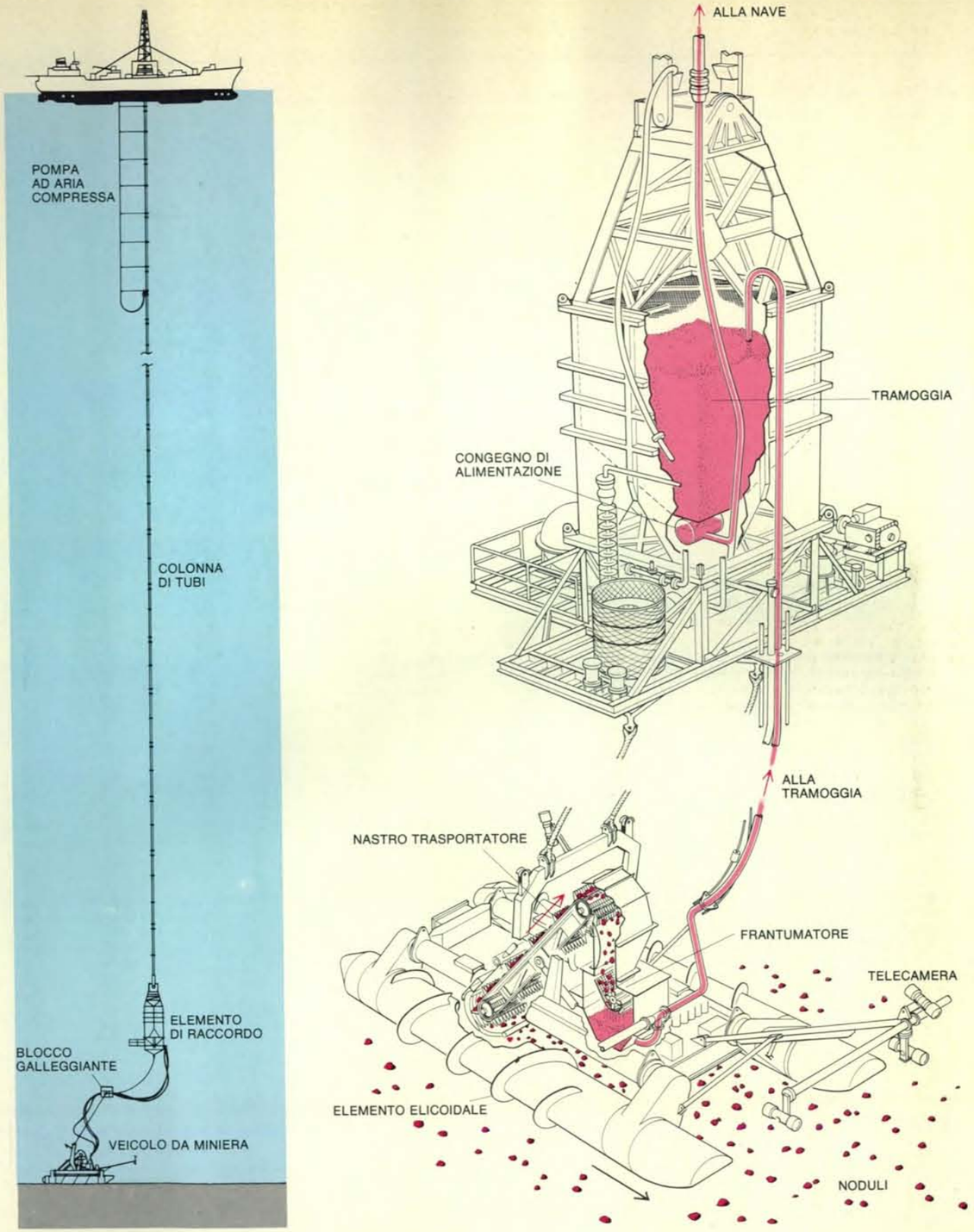
Nel primo Comitato, a cui era stato assegnato il compito di stabilire il regime internazionale per la gestione del patrimonio comune del genere umano, gli schieramenti politici erano un po' diversi, in quanto l'agente polarizzatore era chiaramente la divergenza di interessi fra le nazioni industrializzate del Nord e quelle in via di sviluppo del Sud. Il Sud sosteneva che l'adozione, avvenuta nel 1970, della Dichiarazione di princìpi da parte dell'Assemblea generale delle Nazioni Unite aveva dato al patrimonio comune del genere umano la sanzione del diritto internazionale. A suo modo di vedere, il compito del primo Comitato era quello di creare un forte ente per i fondali marini, la Seabed Authority, con ampie funzioni e vasti poteri. Tale ente avrebbe avuto un braccio operativo, l'Enterprise, che si sarebbe occupato direttamente delle ricerche e dello sfruttamento del patrimonio comune per conto di tutto il genere umano e avrebbe potuto costringere le società minerarie a vendergli la tecnologia necessaria per quella funzione. Il Nord reagì a questa proposta proponendo a sua volta un ente autorizzato a concedere licenze e a riscuotere una piccola tassa, o royalty, sullo sfruttamento del patrimonio comune, che per il resto poteva essere utilizzato dal Nord a suo piacimento.

Il compromesso che risolse questa controversia fu lungo e difficile da raggiungere. Il Nord, tranquillo nel suo possesso della tecnologia e dei capitali necessari, definì privo di senso pratico il sistema dell'Enterprise concepito dal Sud. D'altra parte l'ente autorizzato a concedere licenze, che veniva proposto dal Nord, era inaccettabile per la maggior parte delle nazioni riunite alla Conferenza, le quali lo consideravano una violazione del principio del patrimonio comune del genere umano. Il compromesso proposto da Henry Kissinger univa in un «sistema parallelo» l'Enterprise privo di senso pratico e l'inaccettabile ente autorizzato a concedere licenze. Per una sorta di incantesimo aritmetico, questa combinazione avrebbe dovuto offrire una soluzione pratica e nel contempo accettabile!

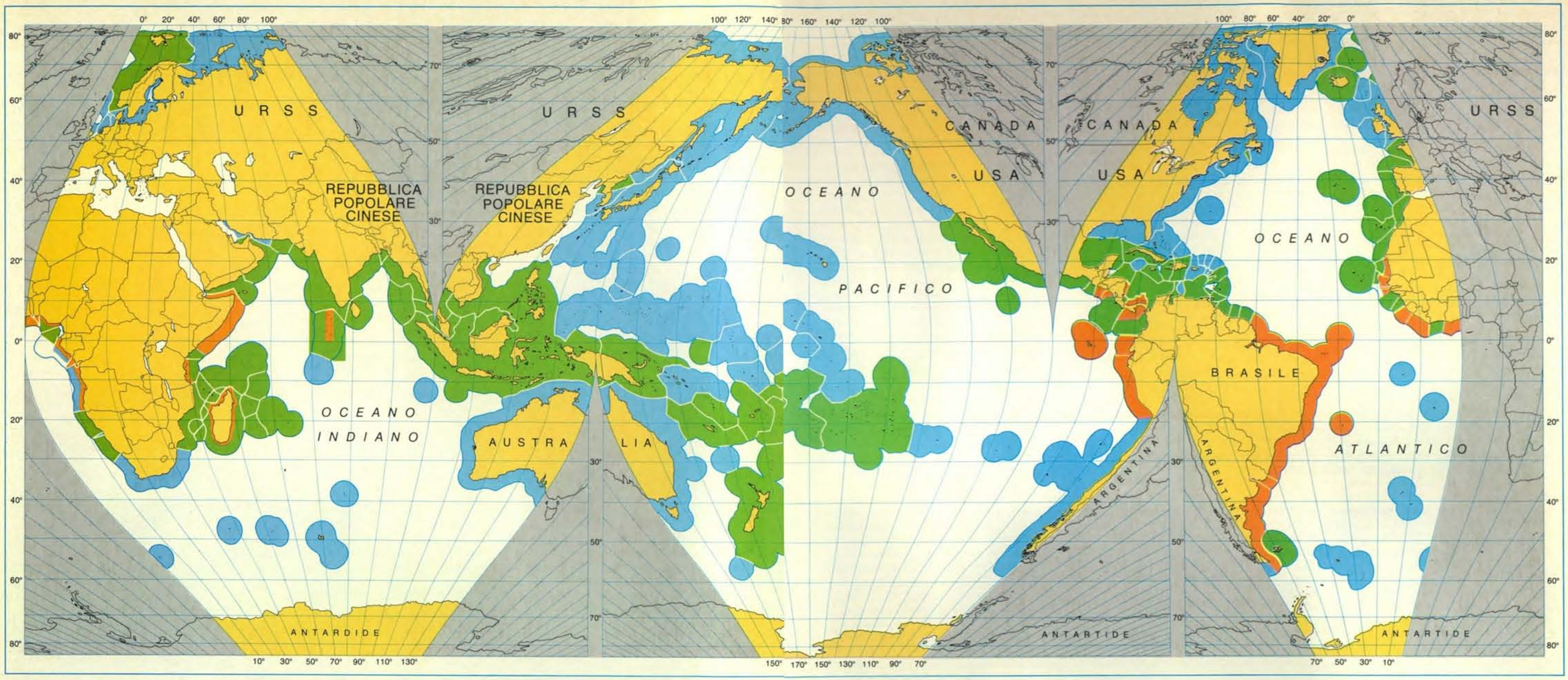
Nel corso delle tediose trattative per giungere a questo compromesso, il testo che definiva il sistema parallelo divenne sempre più astruso e lontano dalla realtà. Facendo leva sul timore di una mancata ratifica da parte del Senato degli Stati Uniti, il negoziatore americano Leigh Ratiner riuscì a far dichiarare esplicitamente in anticipo ogni particolare amministrativo e finanziario e a fargli prendere forma in articoli che non si sarebbero potuti modificare per 25 anni - e questo per un'industria che non esisteva ancora e il cui futuro era dichiaratamente incerto. A questo punto il Sud ebbe qualche ripensamento sulla desiderabilità di sfruttare il patrimonio comune del genere umano. Il calo dei prezzi dei prodotti del suolo privò i paesi in via di sviluppo esportatori di risorse di una nuova fonte di concorrenza per i loro minerali. Contro l'opposizione del Nord (a eccezione del Canada), questi paesi si sentirono più propensi a limitare la produzione dei fondali marini che non a gestirla.

Altri sviluppi hanno contribuito a ridurre ancora di più il valore del patrimonio comune del genere umano, soprattutto dei noduli di manganese. Il Messico e il Cile hanno trovato ricchi giacimenti di noduli nelle proprie zone economiche, la Francia ha dei noduli nelle acque polinesiane e gli Stati Uniti possono rivendicare,






**Sono diversi i sistemi con cui i noduli di manganese, distribuiti su vaste aree del fondo oceanico, potrebbero essere raccolti. Il sistema illustrato è stato ideato da Conrad G. Welling e dai suoi colleghi della Ocean Minerals Company. I noduli contengono tipicamente il 25 per cento di manganese, l'1,5 per cento di nichel, l'1,2 per cento di rame e lo 0,2 per cento di cobalto. In questo sistema un veicolo da miniera *(in basso a destra)* viene spostato sul fondo da due grandi supporti elicoidali. I noduli sono raccolti da un rastrello a nastro continuo, frantumati e pompati sotto forma di impasto liquido in un «elemento di raccordo» *(in alto a destra)*, appeso a una colonna di tubi che fa capo alla nave di superficie.**




**In questa cartina, che si basa su una precedente cartina compilata da F. Webster McBryde della Transemantics, Inc., sono raccolte le rivendicazioni marittime delle varie nazioni. Le aree in azzurro sono le zone di pesca, in cui sono rivendicati i diritti esclusivi di pesca. Le aree in verde sono le zone economiche, in cui vengono rivendicati non solo i diritti di pesca, ma anche i diritti sovrani su tutte le risorse e la giurisdizione sulla ricerca e sull'inquinamento marino. Le aree in arancione sono «acque territoriali», in cui vengono rivendicati pieni diritti sovrani fatto salvo il diritto di «passaggio innocente» (transito senza importanza militare, economica e scientifica). Queste rivendicazioni, che riguardano complessivamente circa il 40 per cento degli oceani, sono riconosciute dalla Convenzione sulla Legge del mare. Il restante 60 per cento è riservato alla tradizionale libertà dei mari, e le risorse in esso presenti sono trasferite con atto legislativo al «patrimonio comune del genere umano». La proiezione cartografica è stata ideata da McBryde per evitare la distorsione sia delle zone di terraferma sia di quelle di mare.**

al largo delle Hawaii e di altre isole del Pacifico, centinaia di migliaia di miglia quadrate di fondale oceanico in cui è probabile che vi siano dei noduli sfruttabili sul piano commerciale. I noduli stessi hanno perso gran parte della loro attrattiva a causa del rinvenimento di depositi di solfuri polimetallici, contenenti rame e altri metalli in concentrazioni molto più elevate, avvenuto per la prima volta in bocche vulcaniche nel *rift* del Pacifico vicino alle isole Galápagos e al largo delle coste dell'Oregon e dello stato di Washington. Depositi analoghi si trovano sicuramente in qualsiasi altro punto delle 40 000 miglia della dorsale tettonica sottomarina, che si estende tutt'intorno al mondo.

Ciò nonostante, la creazione della International Seabed Authority con il consenso praticamente dell'intera comunità mondiale (fra cui vanno contati anche gli Stati Uniti, i quali hanno dato il loro consenso alla Conferenza, e sono sicura che continueranno a darlo) deve essere considerata un grande progresso nelle relazioni internazionali. Si tratta infatti, innanzitutto, di un'istituzione internazionale senza precedenti, in quanto ha il potere di emanare ordinamenti e di agire sulla base del nuovo principio del patrimonio comune del genere umano. In secondo luogo, si tratta di un primo tentativo di politica globale della produzione, che tenga debitamente conto della conservazione dell'ambiente. In terzo luogo, è un'apertura alla cooperazione industriale fra Nord e Sud, basata non sull'assistenza, ma sulla compartecipazione. Cinquant'anni fa la Comunità economica europea era il sogno di alcuni «federalisti». L'International Seabed Authority, un sogno utopistico di 20 anni or sono, è ora una realtà del diritto internazionale. Qualcosa si sta muovendo.

La definizione di un nuovo regime per la ricerca scientifica marina, per la tutela e la conservazione dell'ambiente marino e per l'incoraggiamento del trasferimento di tecnologia è stato il compito del terzo Comitato, il quale ha portato a termine la propria opera prima degli altri in quanto, nelle sue deliberazioni, si è avuta una polarizzazione politica meno accentuata.

La libertà di svolgere ricerche scientifiche, implicita nel concetto di libertà dei mari, è chiaramente limitata nelle acque oceaniche dalla giurisdizione nazionale. Va notato peraltro che, nell'ambito del «regime di consenso» previsto dalla Convenzione, uno stato rivierasco dovrebbe dare il proprio benestare ai cittadini di altri paesi che vogliano svolgere ricerche nelle sue acque. (Gli scienziati americani, che si preoccupavano per i ritardi burocratici nell'ottenere il consenso degli stati rivieraschi, farebbero bene ora a preoccuparsi della mancata adesione del loro paese alla Convenzione, che può impedire loro addirittura di chiedere il benestare, e di ottenerlo.) Il regime del consenso dovrebbe avere l'effetto di incoraggiare la cooperazione bilaterale nelle iniziative di ricerca, adattandole non solo agli interessi dei ricercatori, ma anche a quelli dello stato rivierasco. Più di qualsiasi altro strumento legale del passato, la Convenzione incoraggia l'internazionalizzazione della ricerca marina, rendendo ripetutamente obbligatoria la cooperazione e chiedendo l'intervento delle «competenti organizzazioni internazionali». Si ritiene che eventuali progetti intrapresi sotto il patrocinio di tali organizzazioni debbano ricevere il consenso dello stato rivierasco, a patto che quest'ultimo vi partecipi e non si sia dichiarato contrario al progetto stesso nell'ambito dell'organizzazione.

Le misure per la tutela e la conservazione dell'ambiente marino previste dalla Convenzione riguardano l'inquinamento provocato dai natanti, l'inquinamento prodotto dallo sfruttamento minerario dei fondali marini e dalla produzione di petrolio e l'inquinamento dovuto a fonti situate sulla terraferma e nell'atmosfera. Questa struttura organica per un codice internazionale di leggi prevede perfino la possibile modificazione dell'ambiente con l'introduzione di specie esotiche mediante acquicoltura o l'applicazione di nuove tecnologie. Gli stati sono ritenuti responsabili dei danni provocati non solo in acque sotto la giurisdizione di altri stati, ma anche in alto mare. Le organizzazioni competenti, con particolare rife-

rimento alla Intergovernmental Maritime Organization e all'Environment Programme delle Nazioni Unite, sono tenute a proseguire i lavori in cui sono già impegnate e, in tal modo, vedono rafforzata, la propria autorità.

Anzi, attraverso il suo programma regionale per i mari, l'Environment Programme delle Nazioni Unite ha già dato inizio alla conversione della legge «mite» della Convenzione in una legge «dura», che si possa imporre. Oggi vi sono 11 programmi del genere, che impegnano circa 110 stati e un gran numero di enti intergovernativi e non governativi *(si veda l'illustrazione a pagina 16)*. Le nazioni partecipanti si stanno tassando per coprire il costo di ripulitura dei mari e di prevenzione di ogni ulteriore inquinamento. Mentre la Conferenza era ancora in corso di svolgimento, l'Environment Programme delle Nazioni Unite metteva già in vigore le norme e i criteri universali per il controllo dell'inquinamento, promulgati nella bozza della Convenzione.

Di fronte all'indifferenza, più che alla resistenza, del Nord, i paesi in via di sviluppo hanno fatto propri gli articoli della Convenzione, relativi al trasferimento della tecnologia marina. Essi si sono spinti fin dove hanno potuto, ma il risultato è stato in realtà una legge «mite». Il trasferimento di tecnologia rappresenta un compito complesso dal punto di vista tecnico, sociale, politico ed educativo. Ci si può benissimo chiedere se sarà mai possibile attuarlo in via legislativa. La legge fa quello che può per porre obiettivi e specificare metodi di cooperazione; il resto dipende dalla storia: una storia che si sta già delineando. Nei paesi industrializzati, l'antiquato sistema dei brevetti non ha molta corrispondenza con la realtà della concorrenza tecnologica. La proprietà intellettuale deve proprio trovare il suo posto naturale nel patrimonio comune del genere umano.

La Convenzione dà un importante contributo allo sviluppo di un sistema obbligatorio per comporre le controversie e quindi al mantenimento della pace. Essa stabilisce anzi il sistema globale più organico e più vincolante che sia mai stato concepito. Nel Tribunale internazionale della Legge del mare è stata istituita una nuova corte suprema, che avrà la propria sede, giustamente, nella città anseatica di Amburgo. In una delle sue sessioni speciali, che si occupa appunto dei fondali marini, troveranno ascolto sia gli enti sia le singole persone. Purtroppo (ma la cosa era inevitabile) il sistema offre anche varie scappatoie: gli stati possono chiedere di essere esentati dalle procedure obbligatorie per la soluzione delle vertenze, soprattutto per quel che riguarda le questioni più delicate, come per esempio la delimitazione dei confini delle acque territoriali o le questioni relative agli usi militari del mare.

Sotto questo aspetto, così come da molti altri punti di vista a cui si è fatto cenno, la Convenzione è un documento ambiguo. La Legge del mare non ha comunque un'esistenza autonoma: essa deve essere considerata l'espressione dell'ordine e del disordine del nostro tempo, nazionale oltre che internazionale. Se le attuali tendenze continueranno in altri settori delle relazioni internazionali, gli stati sfrutteranno le scappatoie, le imperfezioni e le contraddizioni della Convenzione per saccheggiare e inquinare il mare, sfruttarne fino all'esaurimento le risorse vive e schierare le proprie armi nelle sue acque profonde. Se, invece, decideranno di avviare seri negoziati di pace e di sviluppo globale equilibrato, essi si accorgeranno che la Convenzione offre un fondamento per costruire un mondo migliore.

Il mare è destinato a svolgere una parte di sempre maggiore rilievo nell'economia mondiale. A giudicare dai primi lievi passi nel campo dell'acquicoltura, la tradizionale tecnologia della caccia e della raccolta lascerà il campo alla coltivazione di piante acquatiche e all'allevamento di animali acquatici. Il fondo marino soddisferà la domanda di metalli molto tempo dopo che l'assottigliamento delle riserve di minerali sulla terraferma avrà esaurito tutte le possibili risorse della tecnologia. Entro la fine del secolo, il fondo del mare fornirà più di metà del petrolio mondiale e l'oceano conterrà energia di riserva nelle maree, nelle onde e nelle correnti, nei gradienti termici e salini, nella biomassa e, non meno importanti in prospettiva, nei rarefatti, ma abbondanti atomi di deuterio delle molecole d'acqua.

La crescente importanza economica del mare, lo inserirà più direttamente nella vita politica delle nazioni e della comunità internazionale. Alcuni paesi stanno creando dei dipartimenti per lo sviluppo dei mari, dei ministeri degli affari marini e altre cose del genere per adempiere a nuove funzioni di gestione delle acque del mare. E stanno aggiornando il proprio diritto marittimo ed estendendo l'applicazione della legge ad aree e attività fin qui non coperte dal punto di vista legale. Armonizzando le proprie norme e le proprie istituzioni con la Convenzione sulla Legge del mare, esse stanno interiorizzando e mettendo in pratica le disposizioni della Convenzione.

Le competenti organizzazioni internazionali tanto spesso evocate nella Convenzione - l'Intergovernmental Maritime Organization, la Food and Agriculture Organization delle Nazioni Unite, l'Intergovernmental Oceanographic Commission dell'UNESCO e l'Environment Programme delle Nazioni Unite - si sono già messe al lavoro. Per quel che riguarda soprattutto i paesi rivieraschi in via di sviluppo, questi organismi hanno il compito di aiutare a dare forma alla legislazione nazionale, a migliorare la gestione delle risorse vive nelle loro zone economiche allargate, a tracciare delle rotte di navigazione e a creare degli schemi per il controllo del traffico, a istituire dei centri regionali per il progresso della scienza e della tecnologia dei mari e ad arbitrare le vertenze.

Anche se la cosa appare paradossale, l'estensione della giurisdizione nazionale al mare non diminuisce la necessità della cooperazione internazionale, ma anzi l'aumenta enormemente. Non ci potrà mai essere cooperazione internazionale senza nazioni ben organizzate. Nella realtà tecnologica ed economica del giorno d'oggi nessun paese potrà mai rimanere sano sul piano economico senza una forte cooperazione internazionale.

Le reti mondiali del programma regionale per i mari stanno aprendo la strada a nuovi modi di cooperazione internazionale. Gli interessi che ogni nazione ha nel mare sono talmente importanti che nemmeno in tempo di guerra turchi e greci, israeliani e arabi, iracheni e iraniani hanno interrotto la cooperazione nei rispettivi programmi. Il successo di un programma regionale per i mari nell'oceano Indiano potrebbe dare impulso in quella regione al concetto di «Mare della pace», che è stato adottato e reiterato dalle Nazioni Unite, ma che non è mai stato rispettato dalle superpotenze. Nel Mediterraneo lo sviluppo di un meccanismo per il reperimento dei fondi necessari per finanziare un analogo programma potrebbe forse far rivivere un'idea avanzata per la prima volta nel 1972 da Malta e dal Messico nelle deliberazioni del Comitato per i fondali marini, cioè l'idea di imporre alle nazioni una tassa sull'uso del mare, sui proventi del petrolio estratto in mare aperto, sulle zone di pesca commerciali e sul commercio marittimo.

La firma della Convenzione a Montego Bay ha fatto entrare in funzione a partire dal mese di marzo una nuova organizzazione, la Commissione preparatoria, incaricata di redigere le norme e le regole per l'International Seabed Authority, l'ente internazionale per i fondali marini, e un ordine del giorno provvisorio per la prima sessione dei consigli di amministrazione di quest'ultima. La convocazione di questa prima sessione deve però attendere la ratifica della Convenzione. Fino a quel momento, che non si sa quanto sia vicino o lontano, la Conferenza ha conferito alla Commissione preparatoria alcuni dei poteri dell'ente stesso. Fra i notevoli poteri, funzioni e responsabilità di cui dispone, la Commissione deve esaminare le domande di esplorazione di un sito fatte da «investitori pionieristici», verificare i dati delle prospezioni, scegliere i siti da riservare al futuro ente per i fondali e al suo braccio operativo, l'Enterprise, e negoziare accordi per il trasferimento di tecnologia all'ente stesso e per l'addestramento di personale proveniente dai paesi in via di sviluppo. In breve, la Commissione preparatoria ha notevoli possibilità operative. Essa offre in certo qual modo un regime interinale per l'esplorazione, la ricerca e lo sviluppo della tecnologia estrattiva dei mari, ed è probabile che alcune attività della Seabed Authority stessa si limiteranno, sino alla fine di questo secolo, all'esplorazione, alla ricerca e allo sviluppo.

Una volta acquisita la tecnologia necessaria - forse attraverso una società a capitale misto o a vari tipi di compartecipazione con compagnie minerarie - la Commissione potrebbe applicarla non solo per compiere ricerche esplorative di noduli nell'area internazionale, ma anche per aiutare i paesi in via di sviluppo, che chiedono formalmente questo tipo di assistenza, a svolgere ricerche minerarie e a sviluppare le risorse delle loro zone economiche nazionali. Non c'è nulla nella Convenzione che impedisca sia alla Seabed Authority, sia alla Commissione preparatoria prima di essa, di rendere questi servizi utili e necessari, a condizione che essi siano richiesti da uno stato rivierasco.

Un altro servizio utile sul piano immediato, che la Commissione potrebbe rendere, sarebbe quello di creare dei parchi marini, zone di cui è «disapprovato lo sfruttamento», e di riservarli a scopi scientifici. Il primo parco di questo genere potrebbe fare da cordone intorno alle bocche termali sottomarine delle isole Galápagos e alle strane colonie di crostacei e di giganteschi organismi tubiformi alimentati in quella zona dalla chemosintesi batterica che produce zolfo anziché ossigeno, come nella più familiare e onnipresente fotosintesi (si veda l'articolo *La cresta della Dorsale pacifica-orientale* di Ken C. Macdonald e Bruce P. Luyendyk in «Le Scienze», n. 155, luglio 1981).

Un terzo servizio utile nell'immediato riguarderebbe il controllo degli armamenti e il disarmo. Il trattato del 1972, che mette al bando le armi nucleari dal fondo del mare, non provvedeva alle ispezioni intese a verificare l'osservanza delle sue norme. L'ispezione veniva lasciata per inadempienza alle superpotenze. La maggior parte dei paesi voleva internazionalizzare il controllo e la sorveglianza, ma fu frustrata dalla mancanza di un'organizzazione internazionale competente. La Commissione preparatoria e la Seabed Authority potrebbero colmare questo vuoto.

Nel 1987, quando si dovrà procedere alla revisione del trattato per il disarmo del fondo marino, può darsi che la Commissione avrà acquisito il necessario *know-how*; le stesse tecnologie acustiche, sismiche, magnetiche ed elettroniche impiegate nell'esplorazione delle risorse potranno essere destinate al compito dell'ispezione. Un emendamento al trattato, che sarebbe accolto favorevolmente da una grande maggioranza dei firmatari, potrebbe affidare questo compito alla Commissione preparatoria o alla Seabed Authority stessa se, a quell'epoca, la Convenzione sarà stata ratificata.

La firma della Convenzione a Montego Bay da parte della stragrande maggioranza della comunità internazionale apre così una nuova fase nella costruzione di un ordine internazionale. Essa offre nuovi strumenti per accrescere lo sviluppo economico, promuovere il disarmo, assicurare la tutela dell'ambiente e creare nuove forme di cooperazione scientifica e industriale fra Nord e Sud. Tutte queste possibilità non esisterebbero senza la nuova Legge del mare.


# *I raccoglitori per il 1983*

*Questi raccoglitori corrispondono ai volumi XXX e XXXI della rivista, e rispettivamente ai fascicoli da gennaio (n. 173) a giugno (n. 178) e da luglio (n. 179) a dicembre (n. 184).*

*Sono ancora disponibili i raccoglitori dal Vol. XXII al XXIX e dei raccoglitori non numerati appositamente approntati per sostituire i raccoglitori esauriti.*

*I raccoglitori si possono richiedere direttamente all'editore usando l'apposita cartolina allegata alla rivista e unendo il relativo importo; gli ordini infatti vengono evasi solo a pagamento avvenuto.*

*I raccoglitori si trovano anche presso i seguenti punti di vendita:*

**BOLOGNA**
*Libreria Parolini*
*Via U. Bassi 14*
**FIRENZE**
*Libreria Marzocco*
*Via de' Martelli 22/R*
**GENOVA**
*Libreria Intern. Di Stefano*
*Via R. Ceccardi 40/R*
**MILANO**
*Le Scienze S.p.A.*
*Via del Lauro 14*
**TORINO**
*Libreria Zanaboni*
*C.so Vittorio Emanuele 41*
**NAPOLI**
*Libreria Guida A.*
*Via Port'Alba 20/21*
**PADOVA**
*Libreria Cortina*
*Via F. Marzolo 4*
**PALERMO**
*Libreria Dante*
*Quattro Canti di Città*
**ROMA**
*Claudio Aranci*
*Viale Europa 319 (EUR)*

Ogni raccoglitore L. 3.600

# La scrittura della civiltà di Harappa

*Una fra le quattro più antiche civiltà del mondo ha una scrittura che ha resistito da sempre a ogni tentativo di decifrazione. Oggi, sulla base della documentazione archeologica, si comincia a interpretarla*

di Walter A. Fairservis, Jr.

Le civiltà più antiche sono quelle di Sumer, dell'Egitto, della Cina e della valle dell'Indo, nella parte nordoccidentale del subcontinente occupato oggi dall'India, dal Pakistan e dal Bangladesh. Gli scritti delle tre civiltà nominate per prime possono essere letti, mentre le iscrizioni della civiltà della valle dell'Indo rimangono in gran parte enigmatiche. Quanto sappiamo di questa civiltà deriva perciò esclusivamente dagli oggetti materiali riportati in luce dagli archeologi ed è curioso che, in una parte del mondo famosa per l'antichità della sua letteratura, neppure i testi più antichi contengano riferimenti validi alla prima grande cultura fiorita in quella regione.

La cultura della civiltà della valle dell'Indo è nota come cultura di Harappa, da una delle sue due grandi città: Harappa e Mohenjo-Daro. Le difficoltà che si pongono a chi voglia decifrare la scrittura di Harappa sembrano dapprima praticamente insuperabili: esse derivano in gran parte dall'uso limitato che facevano della scrittura gli abitanti della regione. I loro «testi» consistono quasi esclusivamente in brevi iscrizioni su sigilli e in graffiti, in quantità altrettanto limitata, su ceramiche. Nessuna iscrizione nota consta di più di 21 segni e i testi medi ne comprendono solo cinque o sei. La brevità delle iscrizioni, unitamente al fatto che molti fra i segni usati sono pittografici, ha sciolto le briglie all'immaginazione di molti studiosi. Sono stati compiuti però tentativi molto seri di stabilire relazioni fra la scrittura di Harappa e quella delle civiltà minoica, cananea, ittita e persino con la peculiare «scrittura» dell'isola di Pasqua. Studi compiuti in anni recenti sulla base di ciò che la documentazione archeologica rivela sulla vita nella civiltà di Harappa hanno consentito di impostare in modo più efficace il problema della decifrazione. Io mi propongo qui di prospettare alcuni fra i risultati ottenuti.

Nei decenni trascorsi dopo la scoperta delle due grandi città della cultura di Harappa, avvenuta negli anni venti, e particolarmente dopo la fine della seconda guerra mondiale, gli archeologi in India e nel Pakistan hanno individuato un migliaio di altri siti appartenenti alla stessa cultura. Questi siti sono dispersi su un grande arco che si estende dall'India occidentale, in prossimità del fiume Narmada, verso nord attraverso i distretti di Gujarat e di Kutch, passando poi per le regioni pakistane del Sind e del Panjab occidentale, fino a pervenire nel Rajasthan indiano e nel Panjab orientale in prossimità di Delhi. Altri insediamenti sono stati riportati in luce lungo la costa del Mare Arabico quasi fino al confine con l'Iran e nel Belucistan; una missione archeologica francese ha scoperto di recente un insediamento addirittura in prossimità del fiume Oxo (Amu Darya), nel cuore dell'Asia centrale.

La maggior parte dei siti della cultura di Harappa sono poco estesi, avendo una superficie compresa fra 0,8 e 2 ettari, e si trovano in vicinanza di fiumi o corsi d'acqua. Pare che ci siano state tre fasi distinte di insediamento. I siti della fase più antica si trovano nei territori di confine fra il subcontinente e l'Iran o nelle aree immediatamente circostanti. Quelli della fase matura sono molto più dispersi: per la maggior parte si trovano nella valle dell'Indo, ma includono anche siti lontani, come quelli del Belucistan e quello in prossimità dell'Oxo. Anche gli oggetti riportati in luce in questi siti più tardivi offrono indizi che l'antica cultura di Harappa si stava fondendo con popolazioni pastorali e agricole locali. Queste scoperte suggeriscono che la civiltà di Harappa non abbia avuto termine in modo brusco, ma abbia svolto una funzione nello sviluppo dello stile di vita di villaggio, che è così tipico del subcontinente oggi.

Che cosa ci dicono i reperti archeologici sulla vita nei siti della cultura di Harappa? In primo luogo, essi ci indicano che la risorsa principale era costituita dall'agricoltura: si coltivavano cereali e ortaggi e l'allevamento dei bovini era il cardine dell'economia del tempo. In secondo luogo, ci rivelano che la cultura materiale di quella popolazione era semplice ma non povera. Gli uomini della civiltà di Harappa usavano rame e bronzo per la fabbricazione di alcuni utensili e armi e a volte lavoravano l'oro e l'argento per produrre quelle perline per collane che avevano un posto molto rilevante fra i loro gioielli. Al tempo stesso, la maggior parte dei loro manufatti erano di legno e osso, di conchiglie, selce e argilla, tutti materiali che abbondavano nella regione. Fra le perline, alcune venivano ottenute con grande abilità da pietre semipreziose più rare, come l'agata, la corniola e il lapislazzuli. (Il sito in prossimità del fiume Oxo era vicino anche ad antiche miniere di lapislazzuli.)

Un altro aspetto sorprendente della vita nella civiltà di Harappa era la grande importanza che aveva in essa la standardizzazione. In architettura, i mattoni erano prodotti secondo formati ben precisi e venivano messi in opera rispettando certi canoni fissi. I sistemi di drenaggio e di eliminazione delle acque di scarico seguivano un modello invariabile. Le abitazioni avevano dimensioni costanti e talune speciali strutture (forse pubbliche) erano ubicate rispetto a quelle private secondo una planimetria immutabile. In questo quadro generale rientrano pesi e misure standard, vasellame dalla forma e dall'ornamentazione sempre uguali e manufatti di aspetto unificato, come mestoli, pesi per telaio e anche carri-giocattolo. Al tempo stesso, taluni aspetti della vita di Harappa, per esempio l'uso di acconciature tipiche e di molteplici braccialetti e collane e anche lo stile delle statuine, sembrano preannunciare la posteriore cultura dell'India.

Tanto basti per un elenco sommario di quanto ci è noto. Ma anche quanto non conosciamo può essere altrettanto significativo. Non ci sono indizi di rivalità fra vari stati, di guerre, di commerci internazionali consistenti, o di re e di corti e dei grandi complessi di templi, che sono così tipici delle altre antiche civiltà del Vecchio Mondo. La documentazione archeologica non ci rivela quasi nulla sulla religione o sull'organizzazione politica e sociale della civiltà di Harappa, benché essa abbia avuto grande importanza per le civiltà che le fecero seguito. I cittadini di Harappa coltivarono il cotone e forse il riso, addomesticarono i polli e forse inventarono il gioco degli scacchi e una delle due grandi fonti di energia non muscolare dell'antichità: il mulino a vento. (L'altra fu la ruota idraulica.)

Quando fiorì questa civiltà enigmatica? Benché su questo punto non manchino controversie, pare che la fase matura si sia estesa press'a poco dal 2200 al 1700 a.C. La maggior parte dei siti furono presumibilmente occupati per non più di 200 anni e danno l'impressione di uno sviluppo effimero caratterizzato da una robusta organizzazione di gruppo e da rapporti regolari fra insediamenti contemporanei.

Il compito di «leggere» le iscrizioni della cultura di Harappa è così difficile che molti studiosi lo giudicano irrealizzabile. La prima difficoltà deriva dal fatto che esse rappresentano la scrittura ignota di una lingua ignota per la quale non si conoscono testi bilingui come la Pietra di Rosetta per l'antico Egitto o il monumento di Bisutun (Behistun) in Iran. Una seconda difficoltà consiste nell'assenza di testi lunghi. L'intero corpus degli scritti consta di circa 4000 sigilli, impronte di sigilli e graffiti su vasellame; inoltre molti fra i

**Questo sigillo della cultura di Harappa, conservato nel Museo nazionale del Pakistan a Karachi, mostra uno zebù. I quattro segni nella parte superiore sono fra quelli che compaiono nella griglia alle pagine 28 e 29: da sinistra a destra, i tre tratti che designano il numerale 3, il quale ha anche il significato di «primo» o «il più ragguardevole»; una figura formata da una foglia di pipal, il fico sacro, e da una freccia, col significato di «madre» o «padrona»; una figura umana con «corna», che indica un personaggio di alto rango; infine, un recipiente con manici, un suffisso onorifico col significato di «elevato» o «superiore». Il sigillo apparteneva, a quanto pare, a un membro di alto rango della società, forse una donna.**

**«Il Signore delle bestie», un famoso sigillo conservato nella collezione archeologica del Museo nazionale dell'India a Nuova Delhi, mostra una figura umana seduta con le gambe incrociate, che ha sul capo un'acconciatura con vistose corna di bufalo indiano. Le «bestie» associate alla figura sono un rinoceronte e un bufalo indiano *(a sinistra)*, un elefante e una tigre *(a destra)* e una capra *(in basso)*, alla cui destra forse ce n'era un'altra, andata perduta quando si ruppe il sigillo. L'iscrizione può essere resa così: «Il Nero, il Bufalo Nero *an-i(l)*, l'Alto, il Signore dei Capi».**

**I graffiti su vasellame, come quelli su questi cocci trovati nel Belucistan, sono una seconda fonte di segni tipici della cultura di Harappa. Essi non sono numerosi come le iscrizioni sui sigilli e le impronte di sigilli, ma indicano che le iscrizioni dovevano essere lette da destra verso sinistra.**

sigilli sono danneggiati e molti graffiti sono interrotti in conseguenza di rotture. Una terza difficoltà deriva dal fatto che la civiltà di Harappa fu non solo geograficamente lontana dalle altre civiltà del suo tempo, ma anche storicamente remota rispetto ai posteriori sviluppi culturali nel subcontinente indiano.

Per fortuna, i sigilli di Harappa presentano caratteri scritti e immagini, e i motivi figurati possono fornire indizi circa il significato delle iscrizioni. La maggior parte dei sigilli sono pezzi quadrati o rettangolari di steatite, un materiale che si presta molto bene a essere inciso. Sulla parte posteriore essi presentano per lo più un rilievo con un foro attraverso cui doveva passare una cordicella; la parte anteriore presenta invece una figura in negativo e un'iscrizione. Figura e caratteri incisi identificavano presumibilmente il proprietario del sigillo, cosicché, quando il sigillo veniva premuto su argilla molle, l'impronta ottenuta serviva in un modo o nell'altro come riconoscimento del proprietario del sigillo.

La maggioranza dei sigilli di Harappa presenta uno o l'altro di due motivi figurati distinti: il primo è un animale, di solito un toro dalle lunghe corna, ma sono rappresentati anche zebù, bufali indiani, capre, tori dalle corte corna, rinoceronti, tigri, gaviali (i coccodrilli tipici dei fiumi indiani) ed elefanti. Indipendentemente dal tipo di animale, dinanzi a esso era raffigurato un oggetto: quello associato ad animali selvaggi o pericolosi è simile a un piatto; quello raffigurato davanti agli animali domestici è o un cesto o (particolarmente dinanzi al toro dalle lunghe corna) un emblema in forma di «fusto», che compare anche in raffigurazioni da cui risulta che veniva portato in processione.

Alcuni sigilli con animali li raffigurano a gruppi. Un esemplare ben noto presenta in centro una figura antropomorfica, chiamata talvolta il «Signore delle bestie» e seduta con le gambe incrociate; sul capo ha un'acconciatura caratterizzata dalle lunghe corna del bufalo indiano. A destra di questa sono visibili un elefante e una tigre, a sinistra un rinoceronte e un bufalo indiano. Il sigillo è danneggiato, ma in basso, sotto la figura centrale, è visibile una capra che guarda verso l'alto, e immediatamente alla sua destra si notano quelle che potrebbero essere le corna di un'altra figura di capra disposta in modo speculare, con la restante parte del corpo andata perduta. Un sigillo meno elaborato presenta al centro un pipal o fico sacro *(Ficus religiosa)*; dal suo tronco spuntano le teste di due tori dalle lunghe corna.

Un'altra raffigurazione di un «gruppo di animali» appare su un sigillo prismatico. (Alcuni sigilli della cultura di Harappa avevano, infatti, forma prismatica e pochi altri forma cilindrica.) Su una delle sue tre facce sono raffigurati, in fila, un elefante, un rinoceronte e una tigre; una quarta figura è troppo consumata per essere identificabile. Presumibilmente è questa la faccia del sigillo dedicata agli «animali selvaggi». Sulla seconda faccia è rappresentata una seconda processione formata da quattro animali, nessuno dei quali a quanto pare è «selvaggio». Su entrambe le facce, però, al di sopra della processione di animali è raffigurato un quinto animale: un gaviale, che evidentemente fa parte del gruppo degli animali «selvaggi». Sulla faccia del sigillo riservata a questi animali, il gaviale è raffigurato con un pesce distintamente riconoscibile davanti al suo lungo muso e con quello che potrebbe essere un pesce dietro la coda. Sulla seconda faccia non appare alcun pesce riconoscibile.

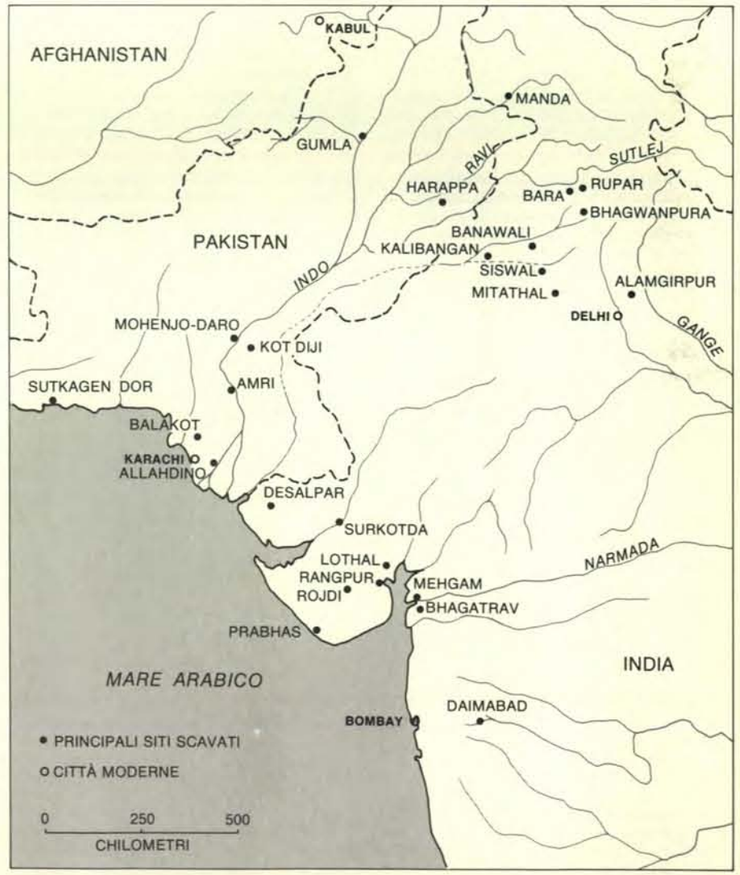
Un quarto sigillo, ancora più elaborato di quello del «Signore delle bestie», introduce un elemento di «venerazione». Esso ripete la figura del fico sacro, questa volta nell'angolo a destra in alto. Fra i rami dell'albero c'è una figura antropomorfica munita di corna. Di fronte a essa si nota una figura inginocchiata e questo personaggio indossa una gonna ed è quindi presumibilmente di sesso femminile; a sinistra della figura inginocchiata c'è una grande capra. Sette personaggi che indossano gonne occupano la metà inferiore del sigillo; i loro capelli sono raccolti in una sorta di lunga «coda di cavallo». Una scena un po' agghiacciante, su un quinto sigillo, presenta varie figure acconciate in modo simile (una delle quali porta una gonna) e attaccate da un bufalo indiano.

Questi motivi danno qualche suggerimento sull'organizzazione sociale della civiltà di Harappa: gli individui che portano i sigilli appartenevano a gruppi che trascendevano i normali lignaggi familiari. Per esempio, tutti gli individui sul cui sigillo appariva un rinoceronte dovevano avere qualche legame sociale in comune, che poteva consistere nell'appartenenza a qualche gruppo sovrafamiliare, come un clan o un'associazione. La raffigurazione su alcuni sigilli di scene di processione, dove pare che effigi di animali venissero portate come stendardi, aggiunge forza a questo concetto di gruppi sovrafamiliari. Questi gruppi, a loro volta, potevano rientrare in un raggruppamento più ampio, costituito da due parti, come viene suggerito dal motivo del «Signore delle bestie» da un lato e da quello dei «veneratori della divinità del fico sacro» dall'altro. Una tale possibile struttura è ben nota agli antropologi nella forma di una società dualistica, caratterizzata dalla classificazione di clan o organizzazioni sussidiarie simili in due gruppi, o «metà», fra i quali avvengono di solito i matrimoni.

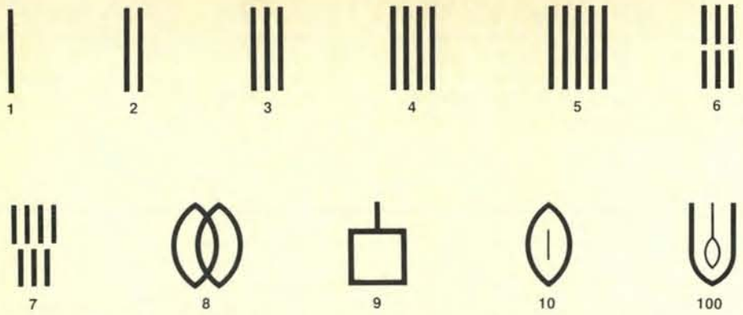
Se si accetta l'ipotesi di lavoro che il materiale figurativo rappresentato su ciascun sigillo serva a identificare il portatore in funzione del clan e della «metà» del gruppo sociale più ampio a cui appartiene, ne consegue logicamente che la parte del sigillo dedicata all'iscrizione potrebbe servire a identificare il portatore del sigillo come individuo. L'iscrizione potrebbe fornire, per esempio, il nome dell'individuo, la sua occupazione, il luogo di residenza, il suo rango o titolo, e informazioni analoghe. A sostegno di una simile interpretazione, lo studio dei testi dei sigilli rivela una considerevole varietà nella sequenza dei singoli segni, e al tempo stesso, una frequente ripetizione di alcuni di essi. Quest'ipotesi fornisce all'ipotetico decifratore una qualche base su cui procedere. Come Michael Ventris sapeva che almeno alcuni fra i testi del Lineare B trovati a Creta erano inventari di oggetti materiali, così chi si accinge a decifrare le iscrizioni della cultura di Harappa può procedere in base al presupposto che si tratti di cercare dei nomi propri ed epiteti elogiativi o altre formulazioni atte all'identificazione dell'individuo.

Già negli anni trenta, lo studioso britannico G. R. Hunter aveva identificato un totale di ben 396 segni distinti. Ricerche più recenti ne hanno aggiunti altri 23, cosicché oggi sappiamo che la scrittura della cultura di Harappa comprende 419 segni. Un'analisi statistica dice anche che essi ricorrono 13 376 volte, in 2290 testi noti. Dei 419 segni, 113 sono presenti solo una volta, 47 compaiono due volte e 59 ricorrono meno di cinque volte. Ciò significa che i restanti 200 segni erano d'uso più o meno generale, e la loro analisi dimostra che, per una buona metà, essi sono combinazioni dell'altra metà.

Queste scoperte dimostrano che la scrittura della civiltà di Harappa non era né alfabetica, come il sanscrito, né logografica (ossia con un carattere per ciascuna parola), come il cinese. Essa si colloca dunque nella categoria nota come logo-sillabica: ciò significa che alcuni segni rappresentano parole, mentre altri hanno un valore puramente sillabico, o di suoni. Altri esempi di scritture appartenenti a questa categoria sono i geroglifici egiziani, gli antichi ideogrammi sumerici e il giapponese moderno. Il fatto che, per metà, i segni di uso comune fossero combinazioni di altri segni usati abitualmente fa pensare che chi scriveva, nella civiltà di Harappa, si servisse di tali combinazioni sia per esprimere idee (come fanno i cinesi quando accoppiano i caratteri per sole e




**I siti della cultura di Harappa sono stati identificati in un'area che si estende a sud, in India, sin quasi a Bombay, a nord fino all'Asia centrale, a est fino a Delhi e a ovest sino al Belucistan. Sulla cartina sono indicati 25 fra i principali siti scavati, comprese le grandi città di questa cultura: Mohenjo-Daro e Harappa. In genere essi sono poco estesi, coprendo una superficie da 0,8 a 2 ettari.**




**I numerali della cultura di Harappa, che in origine facevano probabilmente parte di un sistema numerico in base otto, sono formati da una serie di tratti verticali per i numeri da 1 a 5 e da un sestetto di tratti brevi per indicare il 6. Il segno per il numero 7 è formato da una serie simile di tratti brevi. I segni per 8 e 9 sono rispettivamente pittogrammi di un sole doppio e di un «palo di fondazione». Il segno per il numero 10 è un sole singolo recante un tratto breve al suo interno. Un altro segno identificato, quello per il numerale 100, è un pittogramma di un mortaio con un pestello.**

luna per rappresentare la parola «luminosità») sia per combinare suoni sillabici, allo scopo di indicare la «pronuncia» di una parola.

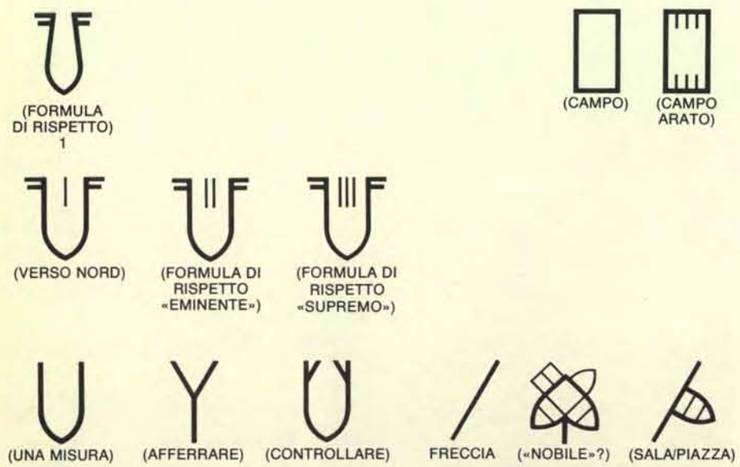
Una parte di cruciale importanza in ogni sistema di scrittura è la serie di mezzi usati per indicare il genere, per distinguere fra singolare e plurale, per stabilire il caso di un verbo e così via. L'identificazione di questi mezzi è molto importante per poter accertare le relazioni che esistono fra i grafemi, o singoli componenti, delle parole e la lingua che i grafemi rappresentano. Ora, i testi della cultura di Harappa presentano un accoppiamento regolare di certi segni, un po' come avviene nelle lingue occidentali per certi titoli: per esempio, «Sua Maestà» o «Vostro Onore». È anche degno di nota il fatto che certi segni appaiano alla metà di un testo ma raramente al principio o alla fine, mentre per altri segni vale l'inverso.

L'ordine abituale dei segni e l'identità degli accoppiamenti caratteristici possono essere stabiliti con l'aiuto di una griglia o reticolato. Poiché le iscrizioni di Harappa sono di solito brevi, tale griglia può constare di un numero abbastanza limitato di colonne verticali: si è di fatto accertato che 14 colonne sono sufficienti *(si veda l'illustrazione alle pagine 28-29)*. Delle prime 17 iscrizioni scelte per essere presentate in righe orizzontali nella griglia, la maggior parte constava di cinque o sei segni; esse furono perciò inserite in caselle corrispondenti alle colonne centrali, lasciando che le righe più lunghe si estendessero verso destra o verso sinistra.

Introducendo i testi nella griglia fu subito evidente che certi segni apparivano regolarmente nella maggior parte delle iscrizioni. Consideriamo i segni «vaso» nella colonna 5, «intreccio da tessitura» nella colonna 8 e «due tratti» nella colonna 10. Quando in un uno stesso testo appaiono due di questi tre segni (come nelle righe *b, e, h* e *p*), essi sono sempre nello stesso ordine da destra a sinistra l'uno rispetto all'altro, indipendentemente da quali altri segni siano inclusi nell'iscrizione. Le loro relazioni sono evidentemente governate da una regolarità di posizione. Perciò, nella costruzione della griglia per i segni della scrittura di Harappa, questi furono posti nelle colonne indicate sopra, anche quando in un testo particolare ne appariva uno solo.

Emerse chiaramente da questa operazione che i segni sopra menzionati erano i più numerosi fra quelli presenti nei 17 testi prescelti. Il segno del vaso, riportato nella colonna, 5, compare 10 volte, e tre sue varianti appaiono in altre due colonne. Il segno dell'intreccio da tessitore nella colonna 8, comprese le sue varianti, si presenta sei volte, e i due brevi tratti verticali della colonna 10, incluse le varianti, ricorrono otto volte. Certi altri segni che utilizzano tratti verticali non hanno posizioni relative altrettanto fisse. Mentre il gruppo costituito da uno o due brevi tratti singoli può essere sistemato nella colonna 10, i gruppi formati da uno, due, tre o più tratti più lunghi non trovano posto nelle colonne centrali della griglia.

A questo punto, dopo essersi familiarizzato con l'aspetto di una cinquantina di segni, quali appaiono nella griglia analitica, il lettore potrebbe chiedersi in quale direzione orizzontale essi debbano essere «letti». Se dovessimo fondarci sulle sole iscrizioni dei sigilli, sarebbe difficile rispondere a tale domanda. Per fortuna, i graffiti incisi sul vasellame hanno fornito una risposta: studi eseguiti da B. B. Lal e da Il Mahadevan, dell'Archaeological Survey of India, hanno dimostrato che alcuni graffiti presentano tratti in parte sovrapposti. La sovrapposizione dimostra che la direzione della scrittura era da destra verso sinistra. Ecco perché la colonna 14 della griglia appare al margine di sinistra e la colonna 1 al margine di destra. L'inversione dell'ordine dei segni permette di disporli nell'ordine di lettura a noi più familiare, cioè da sinistra a destra.




**Le combinazioni di segni costituiscono la metà del totale di 200 segni che ricorrono più spesso nella scrittura di Harappa. Qui sono presentati quattro esempi: a sinistra in alto, uno fra i segni più comuni, un pittogramma di un vaso con manici (indicato col numero arabo 1). Inferiormente a esso sono tre segni combinati, costituiti dal segno indicato col numerale 1 (il vaso con due manici) più segni come quelli per i numerali 1, 2 e 3, ma formati da trattini corti. Sotto ciascun segno si dà il significato in italiano. Sotto i tre segni combinati si osservano, da sinistra a destra, i pittogrammi per un recipiente e per un bastoncino biforcuto, oltre a un terzo segno, che risulta dalla loro combinazione. A destra in alto è il pittogramma per un campo, un semplice rettangolo; accanto a esso, è un segno che combina il pittogramma del campo con i tratti del numerale 6 per formare un nuovo pittogramma, il cui senso sembra essere quello di «campo arato». A destra in basso, da sinistra verso destra, sono i pittogrammi per «freccia» e per «foglia di pipal». Il terzo segno, che combina la «freccia» con una «foglia di pipal» parziale, sembra indicare «sala» o «piazza».**

Veniamo ora alla parte più complessa del problema della decifrazione: qual era la lingua della cultura di Harappa? Quando Ventris identificò la lingua del Lineare B di Creta come greco, superò la maggiore difficoltà. La documentazione archeologica della regione in esame offre qualcosa dello stesso genere in aiuto di coloro che si sforzano di identificare la lingua della cultura di Harappa. Per esempio, i reperti archeologici dimostrano che tale civiltà non si sviluppò in modo subitaneo; essa ebbe una lunga fase preparatoria nel territorio di confine indo-iranico, si diffuse poi su un'area molto estesa e infine dette un suo proprio contributo all'emergere dell'India dei villaggi. Questa sequenza suggerisce che la lingua parlata dalla popolazione della civiltà di Harappa non può essere scomparsa del tutto dal subcontinente.

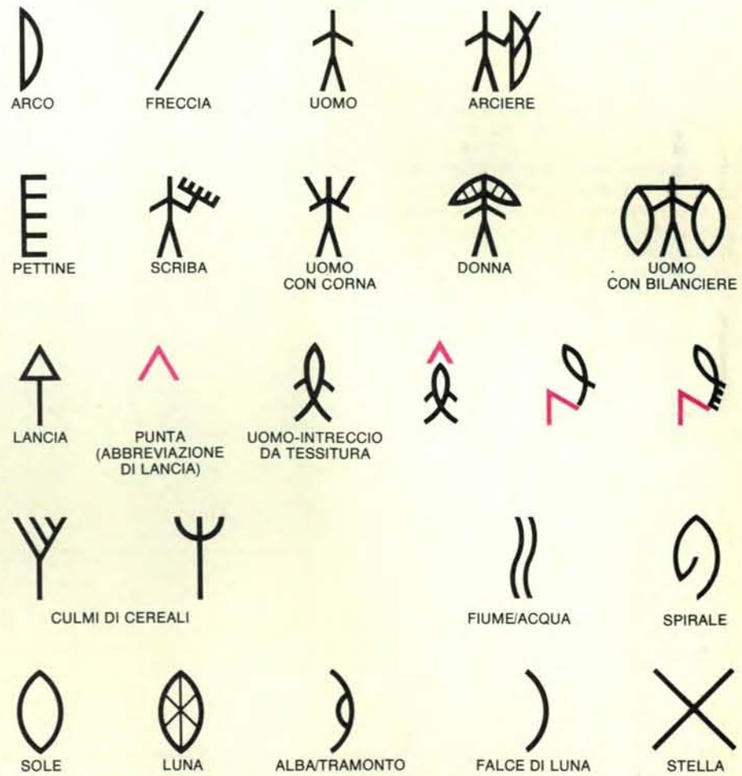
Se accettiamo questo assunto, quale delle tre principali famiglie di lingue parlate nella regione potrebbe essere legata alla civiltà di Harappa? Fra le famiglie linguistiche prese in considerazione vi è quella munda (una famiglia di lingue parlate in gran parte dell'India orientale) che pare presenti affinità con certe lingue dell'Asia sudorientale. Studi sulle forme più antiche di queste lingue trovano però nel loro vocabolario ben poco che si concili con ciò che l'archeologia ci dice sulla cultura di Harappa. Altro possibile candidato è l'indo-europeo, una famiglia di lingue giunte tradizionalmente in India alla metà del II millennio a.C. L'espressione letteraria più antica dell'indo-europeo, il Rigveda, descrive però una cultura fondamentalmente centroasiatica del tutto diversa da quella di Harappa. Il terzo candidato è il dravidico, una lingua parlata oggi per lo più nell'India meridionale e sudorientale, ma che persiste anche in talune sacche dell'India settentrionale e del Belucistan.

Una quarta possibilità è, ovviamente, che la lingua della cultura di Harappa non abbia rapporto con alcuna delle famiglie linguistiche menzionate sopra. Vale nondimeno la pena di considerare la candidatura del dravidico con maggiore attenzione di quella riservata alle famiglie linguistiche munda o indo-europea. Oltre alle forme settentrionali (kurukh, malto e brahui), oggi sono ancora parlate 25 lingue dravidiche, e le famiglie principali (tamil, malayalam, kannada e telegu) sono usate da più di 100 milioni di persone. Grazie, anche in questo caso, alla documentazione archeologica, non è però necessario accettare ciecamente la candidatura del dravidico.

Nel 1974, scavi condotti nel sito di Allahdino, nei pressi di Karachi, riportarono in luce un frammento di avorio dalla sezione semicircolare. Su un lato c'erano fori, a quanto pare praticati per infilarvi bastoncini, sull'altro, c'erano due solchi paralleli longitudinali. L'oggetto era simile a un gran numero di piccole verghe d'avorio riportate in luce negli anni trenta da E. J. H. Mackay a Mohenjo-Daro, la grande città della cultura di Harappa. Gli avori di Mohenjo-Daro non erano - com'è stato asserito - semplici pezzi di un gioco. Per esempio, uno di essi, dalla sezione quadrata, presentava una serie di cerchi e semilune alternati, incisi su una faccia.

Su alcuni avori di Mohenjo-Daro è presente un segno di quelli rappresentati nella griglia (colonna 6, riga *c*), associato a un altro segno che ricorda una pianta. In vari testi di sigilli questi due segni si accompagnano a loro volta a tratti verticali, in numero da uno a sette, e ad altri cinque segni (due dei quali appaiono nella griglia alla colonna 7, riga *c*, e alla colonna 11, riga *q*). Come il secondo segno sugli avori di Mohenjo-Daro, il segno alla colonna 6, riga *c*, sembra rappresentare una pianta, probabilmente il culmo di un cereale. Se si considerano l'associazione dei due segni presenti nelle iscrizioni dei sigilli con quelli che sembrerebbero essere dei numeri e l'incisione su uno degli avori di cerchi e semilune, che si approssimano come numero a un mese lunare (semiluna) di 30 giorni solari (cerchi), i due segni di piante potrebbero logicamente essere interpretati come rappresentanti una parola che significa sia cereale sia mese (o luna). Inoltre, i tratti verticali associati a entrambi i segni non sono mai più di sette, facendo pensare che, per esprimere il numero otto, fosse usato qualche altro segno e, forse, che il sistema numerico della civiltà di Harappa fosse in base otto.

Quale lingua possiede per un cereale, una parola, che significa anche mese o luna ed è associata a un sistema numerico in base otto? Quanto al primo punto, la parola *nel* significa riso in cinque lingue dravidiche, e *nilā* o *nela* significa luna in tre delle stesse cinque lingue e in altre cinque. La medesima parola significa in alcune delle stesse lingue anche mese. Quanto al secondo punto, uno studioso delle lingue dravidiche, Kamir V. Zvelebil, ha fatto notare che il sistema numerico dravidico originario era di fatto, probabilmente, in base otto; il conteggio fino a 10, usato in conformità col sistema numerico attuale in base 10 procede (tradotto) letteralmente così: «uno», «due», «tre», «quattro», «cinque», «sei»,




**Altri pittogrammi riportati nella griglia, riprodotta nelle due pagine seguenti, sono qui presentati assieme ad alcune forme di combinazione. Nella riga in alto, da sinistra a destra, i segni per «arco», «freccia», «uomo» e la combinazione «arciere». Nella seconda riga, «pettine», una terza figura umana formata da semplici linee, che regge un pettine (combinazione che sembra significare «scriba»), una quarta figura umana con «corna», una quinta figura umana con un'acconciatura complessa («donna») e un'altra figura lineare che ha sulle spalle un bilanciere a cui sono appesi due vasi (combinazione che forse indica «guardiano»). Nella terza riga, da sinistra a destra, una «lancia» e l'abbreviazione per «lancia», cioè «punta» *(in colore)*, una formula di rispetto, un «intreccio da tessitura» con braccia umane, una seconda formula di rispetto riservata a chi detiene il governo, e i due segni combinati. Nella quarta riga due culmi di cereali (entrambi i segni hanno il significato intercambiabile di «cereale» o «mese»), un segno col significato di «fiume» o «acqua» e una spirale col significato di «cingere» o «dintorni». Nella riga in basso cinque pittogrammi astronomici: da sinistra verso destra il sole, la luna, l'alba o il tramonto, la falce di luna e una stella.**

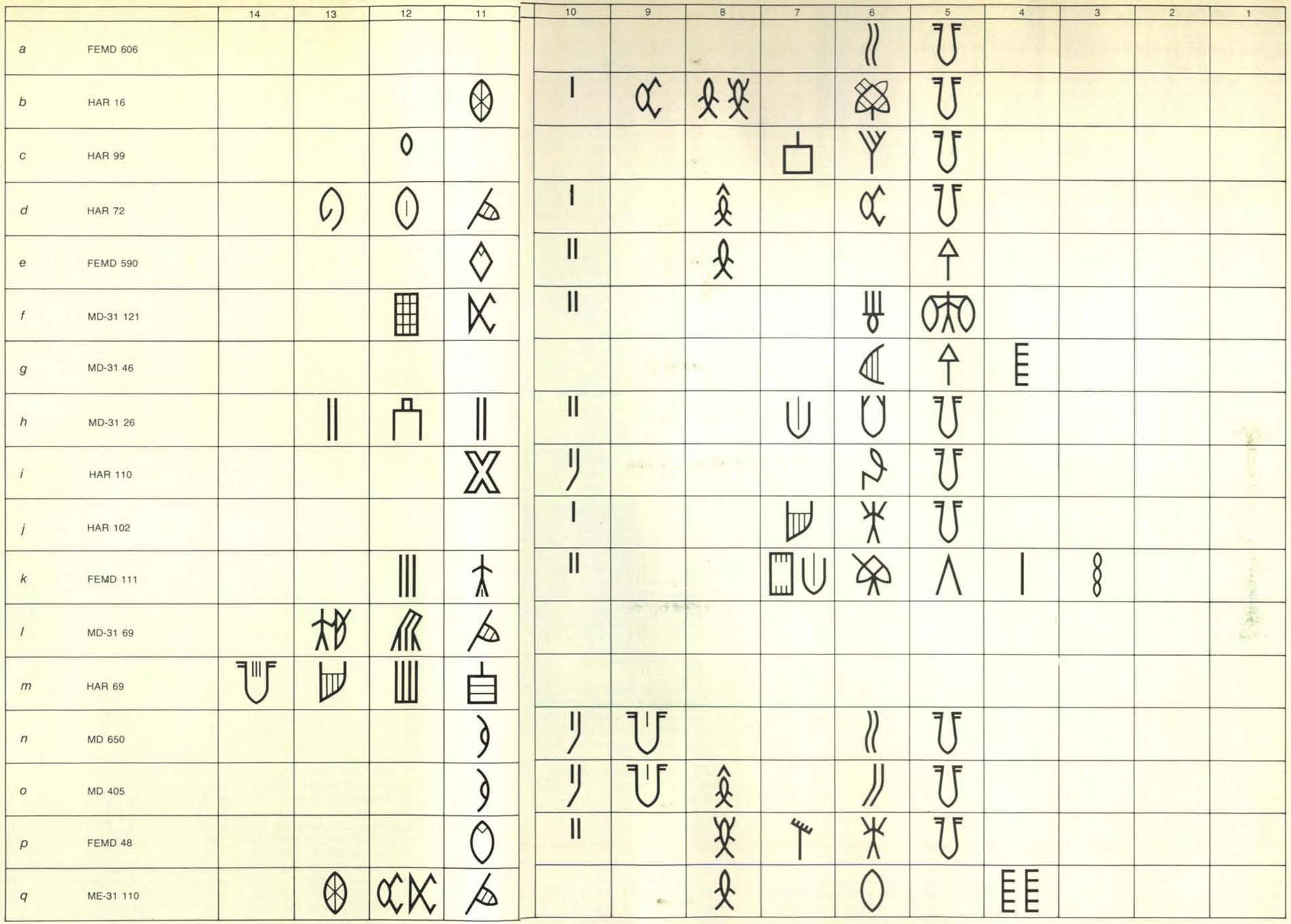
«sette», «numero», «molti meno uno» e «molti».

Nel corso degli anni gli studiosi hanno rilevato la presenza nel dravidico di un certo numero di omòfoni: parole che hanno lo stesso suono ma significati diversi. Per esempio, la parola per designare il bilanciere, la pertica usata per il trasporto di coppie di recipienti appesi alle sue estremità, appoggiando la parte centrale della pertica su uno o su entrambe le spalle è *kā*. La stessa parola significa anche guardiano o protettore. La parola comune per pesce è *mīn*, che è anche la parola per stella. La parola per 100 è *nūṟu*, che significa anche macinare o polverizzare.

Il principio dell'omofonia, o del rebus, si trova in un certo numero di lingue antiche, compresi i geroglifici egiziani. L'elemento chiave del principio dell'omofonia consiste nella ricerca di un equivalente sillabico (per esempio l'immagine di un occhio per significare «io») anziché nell'uso di un segno semplicemente come immagine di una cosa. Per esempio, un antico sovrano egizio, Nr-mr, era rappresentato nei geroglifici dal segno di un pesce gatto *(nr)* e dal segno di uno scalpello *(mr)*. Ciò non significa che egli venisse chiamato «pesce gatto-scalpello», ma solo che il suo nome aveva un suono simile a «pesce gatto-scalpello».

Queste nozioni ci forniscono una base per poter tentare di ricostruire un sillabario della cultura di Harappa? Cercherò ora di verificare quest'ipotesi, avvertendo innanzitutto che la prova del fuoco di qualsiasi tentativo di decifrazione è la sua coerenza interna. Se si stabilisce che un certo simbolo è l'equivalente di un certo suono o di un certo significato, non sarà lecito usarlo poi per rappresentare un qualche altro suono o significato. Una logica interna è essenziale a tutti i sistemi di scrittura; quando i decifratori modificano arbitrariamente i valori dei simboli per adattarli al modello che preferiscono, il loro lavoro è destituito di ogni valore.

Consideriamo innanzitutto il segno rappresentato nella colonna 7, riga *h*. Pittograficamente esso potrebbe essere interpretato come la rappresentazione di un mortaio e di un pestello. Ricostruzioni linguistiche suggeriscono che, nel dravidico, quest'utensile aveva il valore sillabico *nūṟu*. Questo, come abbiamo visto, è anche il suono del verbo macinare o polverizzare e il nome di «100». Consideriamo poi il segno che figura con tanto rilievo nella colonna 5 della griglia. Alcuni anni fa, Mahadevan ipotizzò che questo segno, che appare per lo più alla fine dei testi sui sigilli, fosse una rappresentazione pittografica di un recipiente con manici. Rilevò, inoltre, che varie parole dravidiche per questo tipo di recipiente erano omòfone con parole che significavano maschio, comprendenti anche il comune suffisso onorifico *an* usato con nomi di persona maschili almeno sino all'inizio dell'èra cristiana. Il fatto che un tale segno si trovi alla fine di testi sui sigilli, per i quali è legittima l'ipotesi che contenessero nomi di persona, corrobora la congettura formulata da Mahadevan.

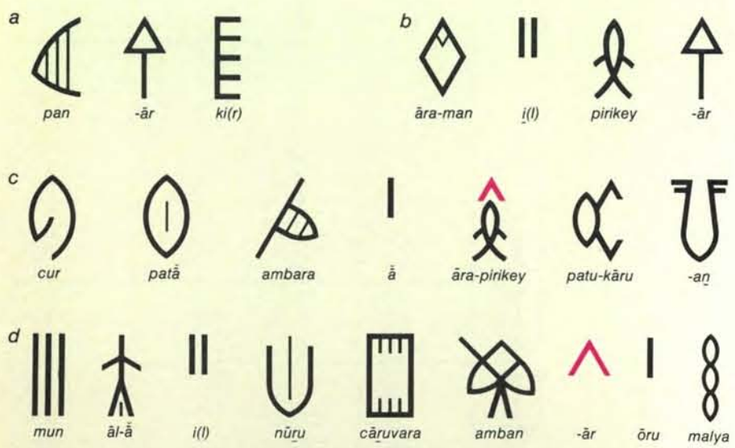



**Questa griglia analitica, comprendente 14 colonne, contiene 17 testi di sigilli, la cui lunghezza varia da due a nove segni. Quando i segni per «vaso» furono assegnati tutti alla colonna 5, i segni «intreccio di tessitura» e le loro varianti alla colonna 8 e i segni composti da due brevi tratti verticali alla colonna 10, fu evidente che se due o più di questi tre segni apparivano in un'iscrizione, essi erano tra loro sempre nello stesso ordine da destra a sinistra, indipendentemente da quali altri segni potessero essere presenti nel testo. Le righe *b, e, h* e *p* forniscono quattro esempi. I singoli tratti verticali nella colonna 10 (righe *b, d* e *j*) rivelano una qualità flessiva. Lo stesso segno appare anche in forma combinata nella colonna 12, riga *d*, in aggiunta al segno per «sole»; nella colonna 11, riga *k*, in aggiunta al segno «uomo», e nella colonna 9, righe *n* e *o*, in aggiunta al segno «vaso». Il senso del singolo segno e della sua combinazione (righe *i, n* e *o*) pare sia quello di un possessivo; il senso del doppio tratto pare sia quello di un locativo. Due testi, alle righe *l* e *m*, non rientrano nella «normale» distribuzione al centro della griglia: si può quindi supporre che siano testi dalla formulazione insolita. Quando, all'interno di una casella della griglia, si trovano due segni, ciò indica che essi sono regolarmente accoppiati nei testi dei sigilli. Nella figura alla pagina seguente in basso si dà un'interpretazione di quattro di questi 17 testi (righe *d, e, g* e *k*).**

**I contatti della cultura di Harappa con la Mesopotamia sono attestati dalla scoperta, in questa regione, di sigilli della cultura dell'Indo. Quello mostrato qui, con il sigillo a sinistra e l'impronta dello stesso a destra, fu scoperto nel 1975 da McGuire Gibson, dell'Oriental Institute dell'Università di Chicago, fra le rovine di una casa del periodo cassitico a Nippur, un'antica città dell'Iraq.**

Ora, un altro dei segni finali sulla griglia (colonna 5, riga *f*) è un pittogramma di una figura umana che trasporta sulle spalle un bilanciere con un recipiente appeso a ciascuna estremità. Come abbiamo visto, la sillaba per un bilanciere, *kā,* è omòfona con parole dravidiche significanti «fare la guardia» o «proteggere». Inoltre una parola dravidica comune per uomo è *āḷ*. La combinazione di queste due sillabe consente di ricostruire il segno della colonna 5, riga *f,* come parola bisillabica *kā-āḷ*. Nelle lingue dravidiche, per ragioni eufoniche, due vocali consecutive sono separate da una consonante, una *v* o una *y*. La parola così ottenuta, *kāvāḷ,* potrebbe essere tradotta «colui che fa la guardia o protegge», una formulazione di identità personale idonea a essere registrata su un sigillo privato.

Per dare altri due esempi, prima di tentare la lettura di alcuni testi di sigilli riportati nella griglia, due fra i segni «numerici» - il segno nella colonna 12, riga *k,* che significa presumibilmente tre, e un secondo segno formato da quattro tratti verticali, che significa presumibilmente quattro - hanno anch'essi omòfoni plausibili. Il primo omòfono, *mu(n),* è uguale alla parola dravidica che significa «primo, il più ragguardevole» *(mun)*; il secondo omòfono, *nāl,* è uguale alla parola che significa «buono» *(nal)*: due aggettivi che si prestano a specificazioni di rango.




**Quattro iscrizioni di sigilli sono qui riprodotte assieme al valore sillabico di ciascun segno in dravidico, una famiglia di antiche lingue parlate nella regione e che sopravvive ancora oggi. L'iscrizione più breve *(a),* nell'interpretazione fattane dall'autore, è «il segno del cantore». L'iscrizione successiva *(b),* che ripete la formula di rispetto in terza persona rappresentata dal segno per «lancia», *ār,* è interpretata come «appartenente alla nobile casa (del) Pirikeyar». L'iscrizione successiva *(c)* è composta da sette segni e comprende una variante del segno «intreccio da tessitura» dell'iscrizione *b*. Può essere interpretata come «Patukaran, potente (nobile) capo degli insediamenti circostanti». L'ultima iscrizione *(d)* comprende nove segni, fra cui il segno con due tratti *i(l)* dell'iscrizione *b* e un'abbreviazione *(in colore)* del segno per «lancia», presente sia in *a* sia in *b*. L'iscrizione suona così: «Appartenente a Munala, padrona di 100 campi arati, nobile primadonna». A tutt'oggi sono stati letti più di 100 testi di sigilli. La maggior parte delle letture si accorda con l'ipotesi che i segni dei sigilli abbiano la funzione di identificare persone.**

Consideriamo ora due testi di sigilli brevi e due più lunghi fra quelli riportati nella griglia. Il testo designato come Mohenjo-Daro 31 46 (riga *g*) è composto da tre segni, che si presume rappresentino uno strumento musicale, una lancia e un pettine. I valori sillabici dravidici assegnati sono, nello stesso ordine, *pan, ār* e *ki(r),* che tradotti corrispondono alla parola «canto», a una espressione onorifica in terza persona e alla parola «segno»; la parola completa viene letta come *Panār-ki(r)*. La traduzione proposta è «sigillo di Panar» o, letteralmente, «il segno del cantore». Il testo designato come FEMD 590 (riga *e*) è formato da quattro segni. Il primo segno, che ha la figura di un rombo, ha il valore trisillabico *āra-man*; il secondo, costituito da due brevi tratti verticali, ha il valore *i(l)*; il terzo, un «intreccio da tessitura», ha il valore trisillabico *piri(key),* e il quarto è la stessa espressione onorifica *ār* del testo precedente. Questo testo viene letto *Āra-man-i(l) pirikeyār,* e per esso viene proposta la traduzione «appartenente alla nobile casa (del) Pirikeyar» («pir» e «pirikey» sono capi).

Il terzo testo, Harappa 72 (riga *d*), è composto da sette segni: esso si apre con una spirale *(cur),* a cui seguono un ovale che racchiude un breve tratto verticale *(patā̆),* una foglia di pipal e una freccia combinate assieme *(ambara),* un breve tratto verticale da solo *(ā̆),* un intreccio sormontato da una punta *(āra-pirikey),* chele combinate col segno del sole *(paṭu-kāru)* e, come segno finale, il vaso con manici *(aṇ)*. Si ottiene una frase piuttosto lunga: *Cur patā̆mbara-ā̆ āra-pirikey paṭu-kāran*. La traduzione proposta è «Patukaran, potente (nobile) capo degli insediamenti circostanti». Il quarto testo, FEMD 111 (riga *k*), è una fra le più lunghe iscrizioni di sigilli tradotte ed è formato da nove segni. Per brevità vengono indicate solo le sillabe e la traduzione proposta: *Munālā̆-i(l) nūṟu cāṟuvara amban ārōru malya,* ossia «appartenente a Munala, padrona di 100 campi arati, nobile primadonna».

Finora sono stati assegnati valori sillabici a quasi 100 segni contenuti in sigilli e in iscrizioni su vasellame; per ciascuno sono stati trovati appropriati omòfoni dravidici. Sono state proposte, inoltre, per più di 100 iscrizioni su sigilli, traduzioni come quelle riportate qui e che vanno da espressioni semplici come «recipiente per acqua» (probabilmente un nome proprio) ad ampollosità come «Arasamban, Grande Capo (dei) Capi del Sudovest, stirpe della Luna».

Ciò che tali testi dimostrano è che un certo numero di individui della civiltà di Harappa (fra cui Arasamban) facevano risalire la loro stirpe a figure importanti del cosmo, come il sole, la luna e le stelle e forse anche la pioggia monsonica. Queste potrebbero essere distinzioni all'interno di ciascun clan. I capi erano associati ad *aramaṇi,* case di capi, che forse avevano funzioni sia residenziali sia amministrative. C'era inoltre un luogo elevato (le «cittadelle» identificate dagli archeologi?) che assolveva una funzione speciale, ma finora non identificata. Un'area destinata a riunioni - un cortile aperto o una sala con colonnato (esempi di entrambi questi spazi sono stati riportati in luce dagli archeologi) - permette di ipotizzare che un'assemblea di capi fosse elemento essenziale dell'organizzazione politica di Harappa.

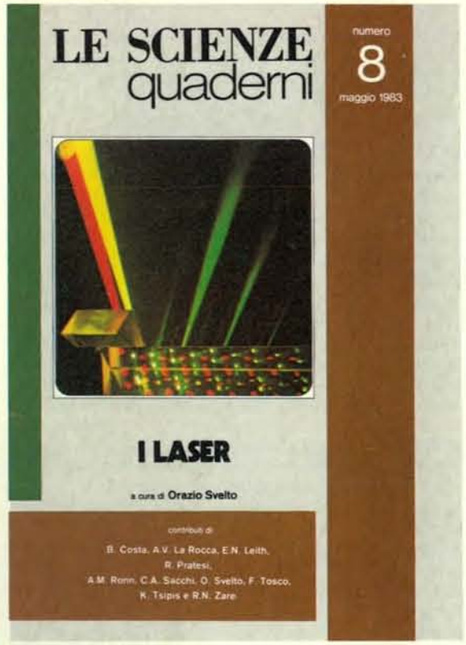
Fra gli altri tipi di capi, che conosciamo dai sigilli, c'erano capi di associazioni (corporazioni?), come quelle dei ramai, degli ispettori ai magazzini, dei supervisori all'irrigazione, e dei proprietari di terreni, una categoria che, come abbiamo visto, includeva anche donne. Sono invece scarsi i riferimenti religiosi. C'è forse una «divinità con corna», designata come «quello di rame» o «il rosso» e forse una «dea madre», ma per il momento si attendono conferme. Fra le figure minori c'erano suonatori di tamburi e cantori: è probabile che i primi convocassero le assemblee e i secondi intrattenessero forse le persone convenute o cantassero durante le cerimonie.

I sigilli attestano inoltre l'esistenza di una classe di scribi, di persone incaricate a sovrintendere ai pesi e alle misure e di ispettori alla distribuzione delle merci, alla macinazione dei cereali e probabilmente a operazioni di caccia. C'erano anche capitani di imbarcazioni e custodi del fuoco. Molti sigilli contengono il segno *kā,* che si riferisce alla funzione di guardia, non in senso militare ma piuttosto nel senso di una responsabilità per la cura dei raccolti e la protezione delle mandrie e delle greggi. Una fra le conferme più incoraggianti che le decifrazioni stanno procedendo nella giusta direzione proviene dal sigillo prismatico in cui si osserva due volte un gaviale al di sopra di una serie di animali di «clan». Nel dravidico la parola per coccodrillo è *mutalai*. La parola dravidica per primo capo, *mutali,* è un omòfono molto simile.

Pare dunque che la lingua di Mohenjo-Daro e di Harappa, circa 4000 anni fa, fosse un'antica lingua dravidica e che gli scribi della cultura di Harappa si sforzassero di tradurre la lingua in forma grafica, come modo per identificare l'élite della civiltà della valle dell'Indo. La civiltà di Harappa era però così diffusa geograficamente che nella lingua deve essere entrato un certo numero di parole non dravidiche, esattamente come parole dravidiche furono in seguito prese a prestito da popolazioni che parlavano l'indo-europeo.

Una grande mole di lavoro di decifrazione rimane ancora da compiere e senza dubbio si troverà che quanto è stato esposto qui non è senza pecche. Il compito che attende i decifratori è ancora più affascinante, poiché dovrebbe condurre a un risultato che fino a poco tempo fa sembrava irraggiungibile: fornire informazioni coerenti sulla forma statale, sull'organizzazione sociale e sull'ideologia della civiltà di Harappa, e forse anche prove più attendibili del fatto che la cultura di Harappa, essendo all'origine dell'India dei villaggi di oggi, non è mai del tutto scomparsa.


Questo mese potete trovare in edicola e in libreria l'ottavo quaderno di «Le Scienze» dal titolo I LASER. Per l'occasione Orazio Svelto, esperto a livello mondiale del settore, ha integrato con numerosi, aggiornatissimi interventi questa raccolta di articoli, illustrando tutte le possibilità offerte da questa nuova, promettente tecnologia.




Otto QUADERNI all'anno, ogni mese da ottobre a maggio. Formato cm 21 × 29. Prezzo di copertina: L. 4000

In questo numero:
*Applicazioni biomediche del laser* di C. A. Sacchi e R. Pratesi
*Il laser nei processi lavorativi* di A. V. La Rocca
*La separazione laser degli isotopi* di R. N. Zare
*Ologrammi in luce bianca* di E. N. Leith
*Armi laser* di K. Tsipis

inoltre di O. Svelto:
*I principi, la storia e gli sviluppi dei laser*
*Le applicazioni dell'olografia*
*Metrologia laser*
*Le applicazioni militari dei laser*
*Lettura e registrazione d'informazione mediante laser*

e di B. Costa e F. Tosco:
*Comunicazioni su fibre ottiche*

Non dimenticate l'appuntamento a ottobre con il quaderno:
GLI OROLOGI BIOLOGICI a cura di Alberto Oliverio.

# La microprogrammazione

*Nei calcolatori moderni il flusso delle informazioni è controllato, al livello più basso, da un microprogramma: un insieme di istruzioni memorizzate che sta al posto di un sistema di controllo cablato*

di David A. Patterson

Come controllare, o pilotare, i segnali elettrici che rappresentano le informazioni costituisce un problema fondamentale nella progettazione di qualunque calcolatore. Nell'unità aritmetica e logica, dove avviene l'effettiva elaborazione dell'informazione, i segnali debbono essere instradati fra vari contatori, sommatori e altri componenti. Il sistema di controllo deve anche mediare il trasferimento di informazioni fra l'elaboratore centrale, le unità della memoria principale e i vari dispositivi di ingresso e di uscita. In uno dei modi possibili di risolvere il problema, il sistema di controllo è completamente cablato, cioè definito in forma permanente nei circuiti elettrici dell'elaboratore. Esiste tuttavia un'altra impostazione, più flessibile e in molti casi meno costosa. L'idea è in sostanza quella di ridurre la complessità del sistema di controllo registrando, in forma codificata, le istruzioni particolareggiate necessarie per controllare il calcolatore. In altre parole, la successione di percorsi che un segnale deve seguire è depositata in un programma, che viene immagazzinato in una unità di memoria distinta, incorporata nell'elaboratore.

Nella gerarchia dei programmi che consentono il funzionamento di un calcolatore, le istruzioni eseguite dal sistema di controllo occupano il livello più basso, il più elementare: ogni istruzione specifica un unico, particolare stato funzionale della macchina. Dato che le istruzioni di controllo riguardano particolari così di dettaglio, il compito di definirle e di codificarle viene chiamato microprogrammazione, introducendo una distinzione fra questa attività e la stesura di programmi di alto livello, per i quali si parla genericamente di *software*. Si dice che un insieme di istruzioni di controllo - un microprogramma - è scritto in microcodice. L'idea della microprogrammazione non è affatto nuova, ma risale a oltre 30 anni fa, all'epoca immediatamente successiva alla realizzazione dei primi calcolatori. A quel tempo l'*hardware* necessario per tradurre l'idea in realtà non esisteva ancora, ma oggi il metodo è stato adottato nella maggior parte dei calcolatori.

Negli ultimi anni si è fatta non poca confusione intorno al significato del termine «microprogrammazione», in gran parte per l'introduzione del microelaboratore, il «calcolatore su un singolo chip», che costituisce il cuore dei più recenti prodotti della progressiva miniaturizzazione nella tecnologia dei semiconduttori, basata sul silicio. Bisogna avere ben chiaro che la microprogrammazione di un calcolatore non si identifica con la programmazione di un microcalcolatore: in linea di principio qualunque calcolatore, dai massimi sistemi «*mainframe*» fino ai più piccoli «*personal computer*», può essere progettato con un sistema di controllo microprogrammato. Per evitare questa confusione a volte i microprogrammi sono classificati come «*firmware*», con un termine che in qualche modo indica anche il loro stato a mezza via fra hardware e software. Per amor di chiarezza, in questo articolo userò questa convenzione: chiamerò macroistruzione una istruzione di software e parlerò invece di microistruzioni quando vorrò fare riferimento a istruzioni di firmware.

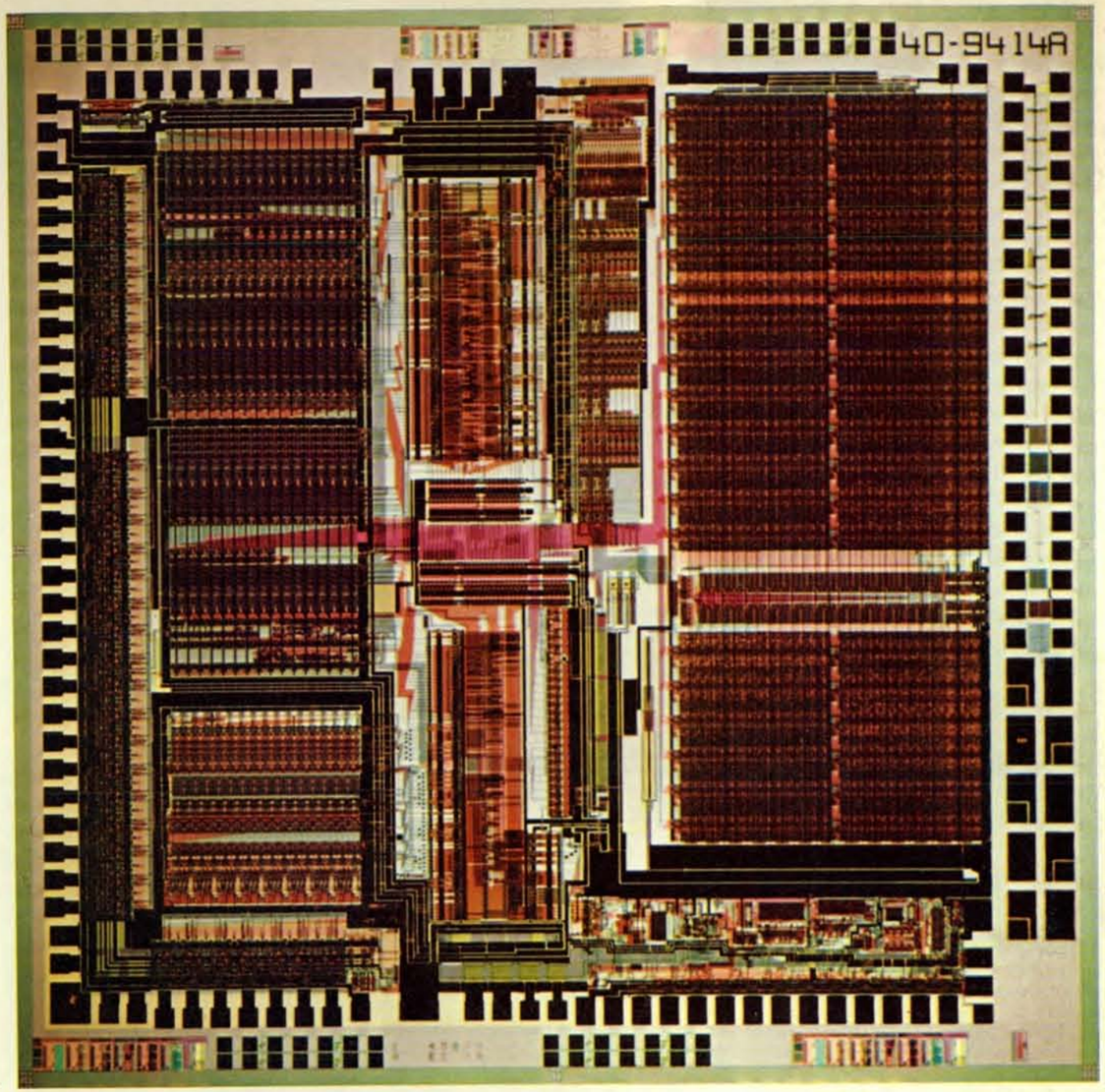
Prima di poter spiegare la microprogrammazione, è necessario passare in rassegna alcuni principi di base dell'hardware e della programmazione dei calcolatori. Sotto molti aspetti un calcolatore non differisce da una calcolatrice tascabile. In ambedue gli apparecchi è possibile introdurre dei numeri, per mezzo di una tastiera, ed è possibile leggerli su una unità di visualizzazione. Alcune calcolatrici dispongono di un tasto di memoria, che consente la registrazione di un numero, in modo da poterlo in seguito recuperare; anche l'unità di memoria in un calcolatore ha il compito di archiviare numeri e di consentirne in seguito il recupero. La memoria di un calcolatore, però, è in grado di gestire milioni di numeri e, per poter conservare traccia di così tanti elementi, a ciascuno viene assegnato uno specifico indirizzo.

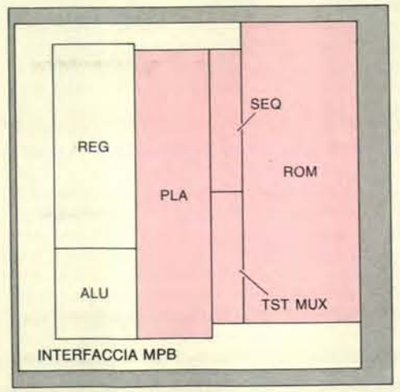
Sia a una calcolatrice tascabile, sia a un calcolatore, bisogna dire esattamente che cosa devono fare. Una istruzione è per un calcolatore esattamente quel che è, per una calcolatrice, la pressione di un tasto di funzione. Per esempio, quasi tutti i calcolatori accettano le macroistruzioni ADD, SUB, MULT e DIV, che corrispondono esattamente ai tasti di una calcolatrice contrassegnati dai simboli +, −, × e ÷. Analogamente, la macroistruzione LOAD (che significa letteralmente «carica») svolge lo stesso ruolo del tasto della memoria in una calcolatrice.

A differenza di una semplice calcolatrice, però, un calcolatore è eminentemente una macchina che prende decisioni. Per capire la distinzione, consideriamo la sequenza di operazioni rappresentata dalle ultime righe del modulo 1040, il modulo per la dichiarazione dei redditi per le persone fisiche negli Stati Uniti *(si veda l'illustrazione in alto a pagina 35)*. Un programmatore che si trovi di fronte al compito di completare il modulo può cominciare con la stesura di un diagramma di flusso, che gli serve come una sorta di «scaletta» per la costruzione di un programma *(si veda l'illustrazione in basso a sinistra a pagina 35)*. Il contribuente che lavora con una semplice calcolatrice tascabile potrebbe seguire il diagramma di flusso, eseguendo le istruzioni in successione, fino a quando non raggiunge la casella di forma romboidale, forma che indica come in quel punto debba essere presa una decisione. Allora confronterebbe i numeri nelle due righe precedenti: a seconda di quale dei due risulta più grande, sceglierà la strada opportuna per proseguire i suoi calcoli.

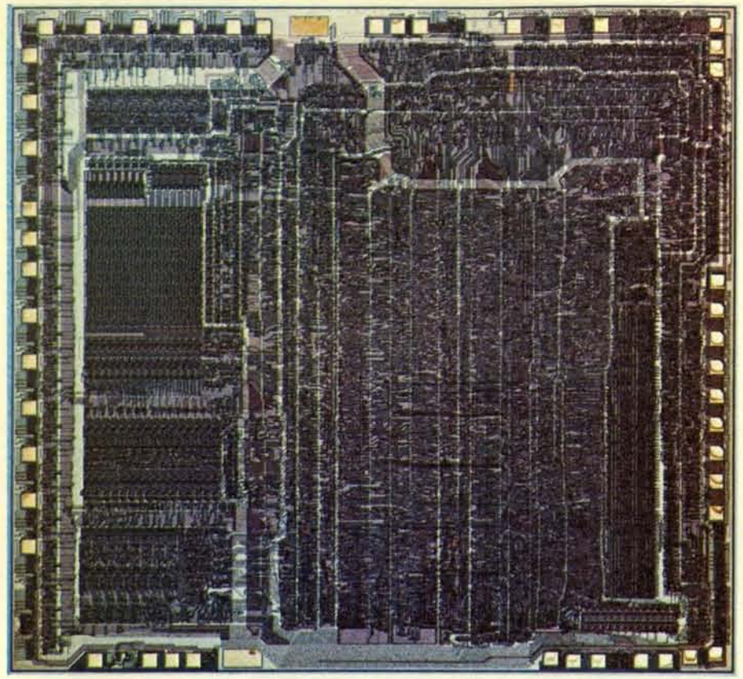
Un contribuente che abbia a disposizione un calcolatore, anziché una semplice calcolatrice, potrebbe fare un passo di più: potrebbe fare in modo che il calcolatore prenda la decisione da solo. Seguendo il diagramma di flusso, in primo luogo direbbe al calcolatore di sommare successivamente i numeri scritti nelle righe della dichiarazione, dalla 50 alla 58 comprese, memorizzando il risultato di ciascuna






**Questo microelaboratore microprogrammato è l'elemento centrale dei calcolatori della famiglia HP 9000 della Hewlett Packard. Il circuito microelettronico ha la densità massima finora raggiunta: contiene l'equivalente di circa 450 000 transistori su un chip quadrato di silicio di 6,3 millimetri di lato. La spaziatura tipica fra le linee che portano i segnali è di un micrometro. Il diagramma a sinistra indica le sette sottosezioni principali del chip. L'area in colore identifica le sottosezioni che costituiscono il sistema di controllo dell'elaboratore e l'area bianca corrisponde al percorso dei dati. Il sistema di controllo è costituito da quattro parti: una memoria a sola lettura *(ROM)*, dove sono immagazzinate le microistruzioni per il controllo del flusso di informazioni in tutto il calcolatore; una matrice a logica programmata *(PLA)*, dove la maggior parte delle microistruzioni vengono decodificate prima di venire trasmesse alle loro varie destinazioni; un dispositivo sequenziatore *(SEQ)* dove viene stabilito l'ordine delle microistruzioni, e un selettore *(multiplexer)* di controllo *(TST MUX)*, dove vengono decodificate talune microistruzioni condizionali. Le tre altre sottosezioni sono una schiera di registri di memoria *(REG)*, dove i dati manipolati dal calcolatore sono immagazzinati sotto forma di «parole», ciascuna della lunghezza di 32 bit; una unità aritmetica e logica *(ALU)*, dove viene effettuata la vera elaborazione delle parole da 32 bit, e un'interfaccia verso il bus memoria-elaboratore *(MPB)* che effettua la supervisione delle comunicazioni fra il chip elaboratore e il resto del calcolatore. Le forme rettangolari sul bordo in alto, in basso e a destra del chip comprendono dispositivi di controllo, schemi di risoluzione speciali e segni di allineamento che non hanno a che fare con il funzionamento del microelaboratore.**



**Questo microelaboratore non microprogrammato si distingue per la sua sezione di controllo cablata, estremamente irregolare, che occupa l'ampia sezione a destra nella fotografia. Le due aree più piccole e più regolari a sinistra corrispondono alle sottosezioni *REG* e *ALU* dell'illustrazione della pagina precedente. Per la sua irregolarità, un sistema di controllo cablato di questo tipo è difficile da progettare e modificare. Quando questo chip fu costruito (nel 1978) i progettisti pensavano che la microprogrammazione avrebbe richiesto un chip di dimensioni molto superiori.**

addizione in uno specifico indirizzo della memoria del calcolatore: in questo caso gli indirizzi dei numeri sono indicati come *L*50, *L*51, *L*52, e via di seguito, e vengono sommati progressivamente a un indirizzo che funge da «accumulatore», chiamato *A (si veda l'illustrazione in basso a destra a pagina 35)*. Il totale di questa successione di passi verrebbe poi immagazzinato all'indirizzo *L*59. Poi il programmatore farebbe lo stesso per le righe dalla 60 alla 66, immagazzinando il totale per questa seconda serie di somme all'indirizzo *L*67. A questo punto (passo 19), darebbe al calcolatore la macroistruzione: «JUMPIF *A* < *L*59, 23». Questa macroistruzione, del tipo chiamato «salto condizionato», dice al calcolatore di cominciare una nuova serie di macroistruzioni, al passo 23, se il numero nell'indirizzo dell'accumulatore è minore del numero immagazzinato all'indirizzo *L*59. In caso contrario, il calcolatore continuerà con il passo 20.

È la macroistruzione di salto condizionato, corrispondente alla casella romboidale nel diagramma di flusso, che fornisce al calcolatore la capacità di prendere decisioni, purché ovviamente il programmatore sappia o possa rappresentare le decisioni come semplici controlli aritmetici. Il programmatore costruisce un programma architettando una gran quantità di salti condizionati che portano ciascun numero fornito in ingresso alla sua corretta destinazione. In media, su dieci macroistruzioni in un programma di calcolatore, una è una macroistruzione di salto condizionato; molti fra i calcolatori di oggi possono eseguire una macroistruzione in meno di un milionesimo di secondo, e pertanto in un secondo un calcolatore, tipicamente, prende più di 100 000 decisioni di questo genere. Le macroistruzioni per le decisioni sono a loro volta codificate sotto forma di numeri e immagazzinate insieme con gli altri numeri gestiti dal calcolatore nella memoria principale. Il principio del programma memorizzato, la cui invenzione ha costituito una pietra miliare nello sviluppo del moderno calcolatore digitale, rende possibile modificare la funzione di un calcolatore cambiando i contenuti della sua unità di memoria, invece di dover modificare il suo hardware.

Con l'aumento della velocità di calcolo, i calcolatori possono eseguire un maggior numero di macroistruzioni e prendere un maggior numero di decisioni, a parità di tempo; quindi diventa possibile far girare programmi di complessità sempre crescente. In questa situazione è difficile evitare gli errori. I programmatori preferiscono usare termini dalle connotazioni meno negative: li sentirete dire spesso che un programma «ha bisogno di una messa a punto». (In inglese la terminologia è più colorita: si parla infatti di un «*bug*», anziché di un errore, e il procedimento di messa a punto è detto «*debugging*», cioè spulciare, togliere le pulci.) Per capire meglio il senso particolare di questo uso linguistico, prendiamo ancora in considerazione l'illustrazione del modulo per la dichiarazione dei redditi riportata nella pagina a fronte. La riga 68 dice al contribuente che cosa deve fare se la cifra riportata alla riga 67 è maggiore della cifra alla riga 59, e la riga 71 spiega che cosa fare se la cifra nella riga 59 è maggiore di quella nella riga 67, ma in tutto il modulo, e nelle istruzioni che lo accompagnano, non si trova un punto che spieghi che cosa si deve fare se le cifre nelle righe 59 e 67 sono uguali! Chiaramente l'Internal Revenue Service degli Stati Uniti ha dimenticato un piccolo errore nel suo programma. Questo è proprio il tipo di dettaglio che un programmatore deve tener presente quando scrive e mette a punto un programma di calcolatore. La difficoltà del compito aumenta fortemente quando si scende ai dettagli finissimi della scrittura e della messa a punto di un microprogramma.

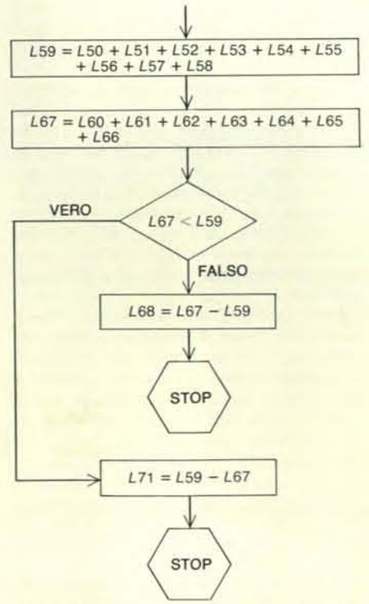
Per vedere come un calcolatore possa prendere decisioni ed eseguire macroistruzioni, dobbiamo guardare un po' più a fondo nel funzionamento dell'unità centrale di elaborazione. L'elaboratore è costituito da due parti principali: il sistema di controllo, che prende le decisioni ed emana di conseguenza i comandi alle altre parti dell'hardware, e l'unità aritmetica e logica, il percorso dei dati, in cui vengono effettivamente svolti i calcoli. Al cuore dell'unità aritmetica e logica sta un sommatore, e in linea di principio tutte le funzioni aritmetiche possono essere ridotte a serie ripetitive di passi di addizione. Ogni parte dell'elaboratore ha la propria schiera di dispositivi di memoria a breve termine, chiamati registri.

A questo livello l'organizzazione del calcolatore assomiglia a un impianto idraulico. Possiamo pensare i fili che collegano i vari dispositivi dell'hardware come tubature che incanalano il flusso di informazione. Il sistema di controllo dirige il flusso inviando un segnale di «on» o di «off» per aprire o chiudere ciascuno dei molti dispositivi elettronici, analoghi a valvole, chiamati porte *(si veda l'illustrazione in alto a pagina 36)*. Le porte debbono chiudersi e aprirsi al momento giusto perché l'informazione possa arrivare correttamente a destinazione, e quindi i segnali di controllo debbono essere inviati con grande precisione temporale. Per questo il ritmo di tutto l'elaboratore è sincronizzato con un orologio interno. Ogni macroistruzione corrisponde a una sequenza di cicli di orologio, dove ogni ciclo contrassegna un singolo trasferimento di informazione lungo il percorso dei dati. La funzione del sistema di controllo è quella di fornire i segnali di controllo durante i cicli di orologio giusti.



| | | | | | |
|---|---|---|---|---|---|
| | 50 | **Balance.** Subtract line 49 from line 40 and enter difference (but not less than zero) ▶ | | | 50 |
| **Other Taxes** (Including Advance EIC Payments) | 51 | Self-employment tax *(attach Schedule SE)* | | | 51 |
| | 52 | Minimum tax *(attach Form 4625)* | | | 52 |
| | 53 | Alternative minimum tax *(attach Form 6251)* | | | 53 |
| | 54 | Tax from recapture of investment credit *(attach Form 4255)* | | | 54 |
| | 55 | Social security (FICA) tax on tip income not reported to employer *(attach Form 4137)* | | | 55 |
| | 56 | Uncollected employee FICA and RRTA tax on tips *(from Form W–2)* | | | 56 |
| | 57 | Tax on an IRA *(attach Form 5329)* | | | 57 |
| | 58 | Advance earned income credit (EIC) payments received *(from Form W–2)* | | | 58 |
| 06 | 59 | **Total tax.** Add lines 50 through 58 ■ | | | 59 |
| **Payments** Attach Forms W–2, W–2G, and W–2P to front. | 60 | Total Federal income tax withheld | 60 | | |
| | 61 | 1982 estimated tax payments and amount applied from 1981 return | 61 | | |
| | 62 | Earned income credit. If line 33 is under $10,000, see page 15 of Instructions | 62 | | |
| | 63 | Amount paid with Form 4868 | 63 | | |
| | 64 | Excess FICA and RRTA tax withheld (two or more employers) | 64 | | |
| | 65 | Credit for Federal tax on special fuels and oils *(attach Form 4136)* | 65 | | |
| | 66 | Regulated Investment Company credit *(attach Form 2439)* | 66 | | |
| | 67 | **Total.** Add lines 60 through 66 ▶ | | | 67 |
| **Refund or Amount You Owe** | 68 | If line 67 is larger than line 59, enter amount **OVERPAID** ▶ | | | 68 |
| | 69 | Amount of line 68 to be **REFUNDED TO YOU** ▶ | | | 69 |
| | 70 | Amount of line 68 to be applied to your 1983 estimated tax ▶ | 70 | | |
| | 71 | If line 59 is larger than line 67, enter **AMOUNT YOU OWE.** Attach check or money order for full amount payable to Internal Revenue Service. Write your social security number and "1982 Form 1040" on it. ▶ (Check ▶ ☐ if Form 2210 (2210F) is attached. See page 16 of Instructions.) ▶ $ | | | 71 |

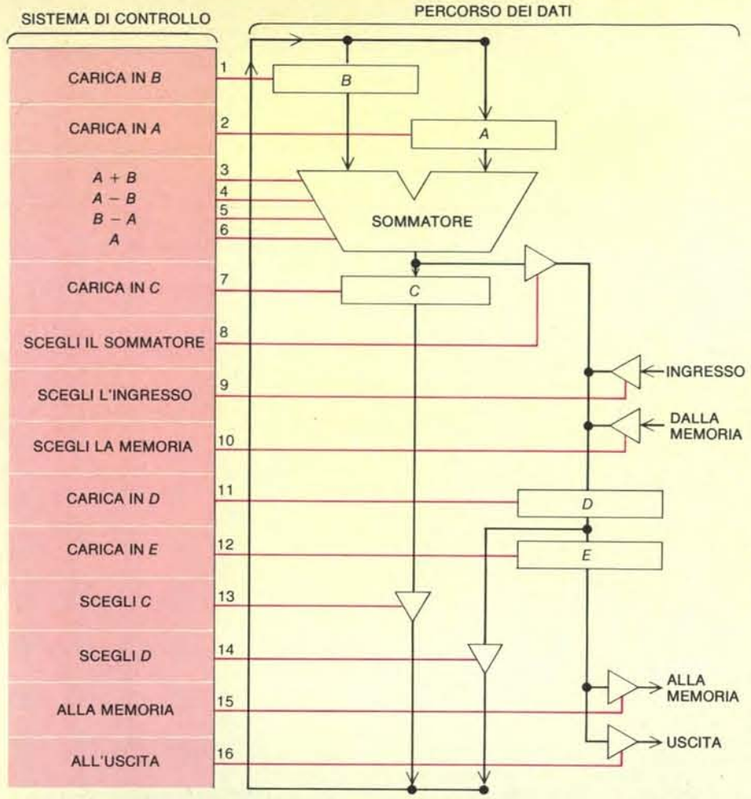
**Questo modulo dà un esempio di una serie di operazioni effettuabili con un programma di calcolatore. L'illustrazione riproduce le ultime 22 righe del modello 1040, la dichiarazione dei redditi delle persone fisiche degli USA per il 1982, preparato dall'Internal Revenue Service.**



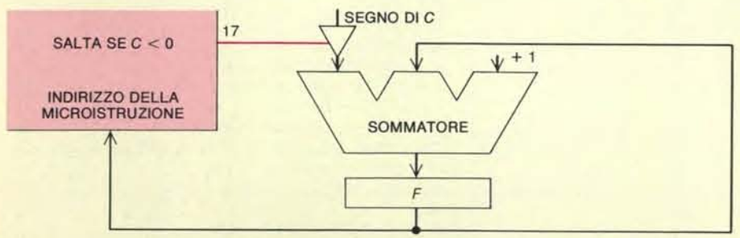

**Questo diagramma di flusso funge da scaletta per la successione delle operazioni necessarie per completare le ultime 22 righe del modulo 1040. Le cifre scritte su ciascuna riga del modulo vengono immagazzinate in un indirizzo particolare della memoria del calcolatore; qui gli indirizzi dei numeri sono indicati come *L*50, *L*51, *L*52 ecc. La casella a forma di rombo è il simbolo usato per un passo «decisionale».**

| INDIRIZZI | ISTRUZIONI | COMMENTI |
|---|---|---|
| 1 | LOAD *L*50 *A* | *A* = *L*50 |
| 2 | ADD *L*51, *A* | *A* = *L*50+*L*51 |
| 3 | ADD *L*52, *A* | *A* = *L*50+*L*51+*L*52 |
| 4 | ADD *L*53, *A* | *A* = *L*50+*L*51+*L*52+*L*53 |
| 5 | ADD *L*54, *A* | *A* = *L*50+*L*51+*L*52+*L*53+*L*54 |
| 6 | ADD *L*55, *A* | *A* =*L*50+*L*51+*L*52+*L*53+*L*54+*L*55 |
| 7 | ADD *L*56, *A* | *A* =*L*50+*L*51+*L*52+*L*53+*L*54+*L*55+*L*56 |
| 8 | ADD *L*57, *A* | *A* =*L*50+*L*51+*L*52+*L*53+*L*54+*L*55+*L*56+*L*57 |
| 9 | ADD *L*58, *A* | *A* =*L*50+*L*51+*L*52+*L*53+*L*54+*L*55+*L*56+*L*57+*L*58 |
| 10 | STORE *A*, *L*59 | *L*59 = *L*50+*L*51+*L*52+*L*53+*L*54+*L*55+*L*56+*L*57+*L*58 |
| 11 | LOAD *L*60, *A* | *A* = *L*60 |
| 12 | ADD *L*61, *A* | *A* = *L*60+*L*61 |
| 13 | ADD *L*62, *A* | *A* = *L*60+*L*61+*L*62 |
| 14 | ADD *L*63, *A* | *A* = *L*60+*L*61+*L*62+*L*63 |
| 15 | ADD *L*64, *A* | *A* = *L*60+*L*61+*L*62+*L*63+*L*64 |
| 16 | ADD *L*65, *A* | *A* = *L*60+*L*61+*L*62+*L*63+*L*64+*L*65 |
| 17 | ADD *L*66, *A* | *A* = *L*60+*L*61+*L*62+*L*63+*L*64+*L*65+*L*66 |
| 18 | STORE *A*, *L*67 | *L*67 = *L*60+*L*61+*L*62+*L*63+*L*64+*L*65+*L*66 |
| 19 | JUMPIF*A*<*L*59, 23 | VAI ALL'INDIRIZZO 23 SE *A* < *L*59 (CIOÈ SE *L*67 < *L*59) |
| 20 | SUB *A*, *L*59 | (*L*67 ⩾ *L*59) PERTANTO *L*67 − *L*59 |
| 21 | STORE *A*, *L*68 | *L*68 = *L*67 − *L*59 |
| 22 | STOP | FINE |
| 23 | LOAD *L*59, *A* | (*L*67 < *L*59) PERTANTO *L*59 − *L*67 |
| 24 | SUB *L*67, *A* | *A* = *L*59 − *L*67 |
| 25 | STORE *A*, *L*71 | *L*71 = *A* (= *L*59 − *L*67) |
| 26 | STOP | FINE |

**Questo programma di macroistruzioni per il completamento del modulo 1040 segue il diagramma di flusso. Ogni macroistruzione è immagazzinata a sua volta in un particolare indirizzo di memoria *(prima colonna a sinistra)*. I numeri agli indirizzi *L*50, *L*51, *L*52 ecc. vengono sommati in successione a un indirizzo «accumulatore» indicato come *A*. I passi da 1 a 10 corrispondono alla casella più in alto nel diagramma di flusso e i passi da 11 a 18 corrispondono alla seconda casella dall'alto. La macroistruzione di salto condizionato al passo 19 equivale alla casella di decisione a forma di rombo. I passi 20 e 21 corrispondono al percorso «falso» del diagramma di flusso, mentre i passi 23, 24 e 25 corrispondono al percorso «vero». I commenti a destra sono note del programmatore per ricordare che cosa sta succedendo; non influenzano l'esecuzione del programma.**




**Questo elaboratore centrale ideale si basa su 16 fili di controllo *(linee in colore)* per guidare i segnali elettrici che rappresentano le informazioni lungo un percorso dei dati semplificato *(linee in nero)*. In questo caso l'unità aritmetica e logica possiede solo un dispositivo dedicato: un sommatore, che riceve numeri da due fili di ingresso, li somma o li sottrae in base a comandi ricevuti attraverso i fili di controllo da 3 a 6, e invia il risultato lungo un filo di uscita. Le forme rettangolari sono registri, componenti di memoria a breve termine che possono contenere una parola di informazione alla volta. Le forme triangolari sono porte, dispositivi che lasciano passare l'informazione quando il filo di controllo associato è attivato e la bloccano quando il filo non è attivato. In un elaboratore non microprogrammato ciascuno dei comandi nella sezione di controllo è definito da un hardware di controllo specifico. Nella versione microprogrammata ogni comando è immagazzinato sotto forma di microistruzione nella memoria di microprogramma. Un microprogramma campione per questo elaboratore è dato nell'illustrazione in alto nella pagina a fronte.**




**È necessario altro hardware perché la versione microprogrammata dell'elaboratore semplificato nell'illustrazione in alto in questa pagina possa prendere decisioni di salto condizionato. Il registro *F* contiene l'indirizzo, nella memoria di microprogramma, della microistruzione che al momento ha il controllo del percorso dei dati. La funzione del sommatore in più è quella di sommare uno all'indirizzo della microistruzione attuale, trasmettendo quindi l'indirizzo della microistruzione che segue nel microprogramma. Il sommatore inoltre rinvia informazioni dal percorso dei dati al sistema di controllo. Il segno del numero nel registro *C* passa attraverso una porta controllata dal filo 17 nel terzo ingresso del sommatore. Quando questo filo è «off», viene eseguita la microistruzione a $F + 1$, l'indirizzo successivo (perché $0 + F + 1$ equivale a $F + 1$). Il sommatore somma o $0 + F + 1$ o $1 + F + 1$ a seconda che il numero nel registro *C* sia positivo o negativo. Quindi con l'aiuto del cavo di controllo 17 il microprogrammatore può effettuare un salto condizionato a una microistruzione, dove la condizione per il salto è il segno del numero in *C*. Il microprogramma per questa operazione è dato nell'illustrazione in basso nella pagina a fronte.**

In un sistema di controllo cablato bisogna progettare una rete di logica elettronica che riconosca ogni macroistruzione nel repertorio del calcolatore. La macroistruzione in sé è una successione di cifre binarie, codificata in una serie di segnali inviati all'elaboratore. Il sistema di controllo pertanto deve in qualche modo trasformare questa serie di segnali nell'insieme di segnali, del tutto diverso, necessario per aprire e chiudere le varie porte nel percorso dei dati. Inoltre, i segnali emessi dal sistema di controllo debbono essere articolati nel tempo in modo da coincidere con gli opportuni cicli di orologio. Ogni macroistruzione deve dar luogo a una diversa successione di segnali di controllo, e pertanto il sistema diventa sempre più complesso, al crescere del numero delle macroistruzioni. Cambiamenti e messe a punto possono essere effettuati esclusivamente attraverso una modificazione del cablaggio.

Agli inizi dell'industria dei calcolatori la progettazione dell'hardware di controllo era un problema di grande rilievo. Al confronto il compito di progettare gli altri componenti dell'hardware era abbastanza facile. Una telescrivente poteva fungere da dispositivo di ingresso e di uscita e la ripetitività delle strutture di memoria e degli elementi circuitali nel percorso dei dati semplificava la loro progettazione e la loro costruzione. L'hardware necessario per contare i cicli di orologio e per emettere i segnali di controllo al momento opportuno rendeva invece il sistema di controllo molto irregolare, il che a sua volta rendeva molto più difficile il compito di progettare questa parte dell'elaboratore centrale.

Nel 1949 Maurice V. Wilkes dell'Università di Cambridge si dedicò alla ricerca di un modo migliore per organizzare la funzione di controllo. Dopo due anni di studio concluse che l'impostazione migliore era quella di pensare il sistema di controllo come una matrice, una griglia rettangolare, in cui ogni riga di quadratini corrispondesse a un ciclo di orologio e ogni colonna fosse associata a un filo di controllo. Per esempio, l'elaboratore semplificato nell'illustrazione qui in alto a sinistra ha 16 fili di controllo, pertanto la corrispondente matrice di controllo ha 16 quadratini di lato (in alto nella pagina a fronte). Per scegliere una successione di operazioni, è sufficiente quindi inserire una opportuna combinazione di cifre binarie nei quadratini: per tutti i fili di controllo che debbono essere *«on»* durante un particolare ciclo di orologio si scrive un 1 nel corrispondente quadratino, mentre per tutti i fili di controllo che devono essere *«off»* si scrive uno 0.

L'hardware equivalente alla matrice di controllo è una struttura di memoria semplice e ripetitiva. Il contenuto di ciascuna cella in una riga determina lo stato della corrispondente linea di controllo per la durata di un ciclo di orologio. La successione di cifre che costituisce una macroistruzione serve soltanto a scegliere la riga opportuna (o la opportuna serie di righe) nella memoria di controllo. In altre parole, la macroistruzione diventa un indirizzo che individua la riga. La progettazione di un sistema di controllo diventa allora più una questione di software che di hardware: la complessità sta nella scelta della giusta successione di 1 e di 0 da immagazzinare in memoria, anziché nello specificare la giusta combinazione di circuiti cablati per generare segnali di controllo.

Wilkes si rese conto della somiglianza che corre fra questo suo metodo e la comune programmazione, tanto che prese a prestito i termini tipici della programmazione per descrivere le sue idee, aggiungendo sempre però il prefisso «micro-», per indicare la natura elementare della funzione di controllo: da qui derivano «microprogrammazione» e tutta la famiglia di termini a quello apparentati. In particolare Wilkes chiamò microistruzione ciascuna riga di quadratini nella matrice di controllo e microprogramma ogni successione di righe che esegue una singola macroistruzione. La struttura di memoria progettata per contenere il microprogramma viene chiamata memoria di microprogramma. L'idea della microprogrammazione ha reso più facile la comprensione della funzione di controllo e, sostituendo i complessi circuiti di controllo con una schiera rispettiva di celle di memoria, ha reso più facile la costruzione dell'hardware; ma quel che più importa è che ha dato al calcolatore una nuova flessibilità: a questo punto il sistema di controllo può essere modificato senza dover riprogettare l'hardware.

Prendiamo in considerazione un microprogramma campione per l'elaboratore semplificato di cui abbiamo parlato prima *(si veda l'illustrazione in alto in questa pagina)*. Supponiamo che l'operazione controllata dal microprogramma sia la sottrazione del numero immagazzinato nel registro di memoria *B* dal numero immagazzinato nel registro accumulatore *A* e quindi l'immagazzinamento del risultato in *A*. Esaminando il percorso dei dati nel diagramma dell'hardware si può vedere che il dispositivo etichettato «sommatore» è in grado non solo di sommare i due numeri, ma anche di sottrarre l'uno dall'altro. Prima di poter riporre il risultato in *A*, però, bisogna tenerlo temporaneamente nel registro *C*: quindi bisogna caricare, e poi scaricare, il registro *C* sul percorso dei dati che riporta ad *A*. L'operazione di sottrazione richiede due microistruzioni. La prima dice al sommatore di sottrarre il numero in *B* dal numero in *A* e di inviare il risultato in *C*; corrispondentemente vengono attivati i fili di controllo 4 («*A–B*») e 7 («CARICA IN *C*»). La microistruzione successiva sposta l'informazione da *C* in *A*, aprendo una porta nel filo che parte da *C* per far passare l'informazione fino ad *A*; quindi debbono essere attivati i fili di controllo 13 («SCEGLI *C*») e 2 («CARICA IN *A*»).

Consideriamo ora un microprogramma un po' più complicato per il medesimo elaboratore. Il secondo microprogramma richiede un processo decisionale simile al salto condizionato che abbiamo visto nel programma per la dichiarazione dei redditi. La decisione si riduce a una scelta fra due

FILI DI CONTROLLO

| MICROISTRUZIONI | 1 CARICA IN B | 2 CARICA IN A | 3 A + B | 4 A − B | 5 B − A | 6 A | 7 CARICA IN C | 8 SCEGLI IL SOMMATORE | 9 SCEGLI L'INGRESSO | 10 SCEGLI LA MEMORIA | 11 CARICA IN D | 12 CARICA IN E | 13 SCEGLI C | 14 SCEGLI D | 15 ALLA MEMORIA | 16 ALL'USCITA | COMMENTI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | CARICA *A* − *B* IN *C*; VAI A 2 |
| 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | CARICA *C* IN *A* |
| 3 | | | | | | | | | | | | | | | | | CONTINUA |

**Questo microprogramma campione per una singola macroistruzione che può essere eseguita dall'elaboratore semplificato dell'illustrazione in alto nella pagina a fronte consiste di una serie di righe in una matrice di controllo, una griglia rettangolare in cui ciascuna riga corrisponde a una microistruzione e ogni colonna è associata a un filo di controllo. Per ogni filo di controllo che deve essere attivato durante un particolare ciclo di orologio si scrive un 1 nel quadrato corrispondente; si scrive invece uno 0 in ogni quadrato che corrisponde a un filo di controllo che non deve essere attivato. (Nel disegno «on» è rappresentato anche dal colore e «off» dal bianco.) L'operazione controllata dal microprogramma è la sottrazione del numero immagazzinato nel registro di memoria *B* dal numero immagazzinato nel registro accumulatore *A* e il successivo immagazzinamento del risultato in *A*. Come nel caso delle macroistruzioni, i commenti sono solo note per il programmatore.**

FILI DI CONTROLLO

| MICROISTRUZIONI | 1 CARICA IN B | 2 CARICA IN A | 3 A + B | 4 A − B | 5 B − A | 6 A | 7 CARICA IN C | 8 SCEGLI LA MEMORIA | 9 IL SOMMATORE | 10 SCEGLI L'INGRESSO | 11 CARICA IN D | 12 CARICA IN E | 13 SCEGLI C | 14 SCEGLI D | 15 ALLA MEMORIA | 16 ALL'USCITA | 17 SALTA SE C < 0 | COMMENTI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | CARICA *A* − *B* IN *C*; VAI A 2 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | SE *C* < 0 VAI A 4 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | CARICA *D* IN *E*; VAI A 4 |
| 4 | | | | | | | | | | | | | | | | | | CONTINUA |

**Questo microprogramma «decisionale» usa l'hardware di controllo potenziato raffigurato nell'illustrazione in basso nella pagina a fronte. L'operazione controllata dal microprogramma è il caricamento nel registro *E* del numero che si trova nel registro *D*. Il salto condizionato richiede l'aggiunta di un quadrato nella matrice, per rappresentare il comando di saltare alla microistruzione 4 se il numero nel registro *C* è negativo (colonna 17). Quindi la macroistruzione di salto condizionato si riduce a una scelta fra la microistruzione 3 e la microistruzione 4.**

microistruzioni. A questo fine bisogna aggiungere un quadrato in più alle microistruzioni e un cavo di controllo in più all'elaboratore. L'hardware di controllo così potenziato è raffigurato nell'illustrazione in basso di pagina 36; il microprogramma riveduto è riportato nell'illustrazione in basso della pagina 37.

L'operazione controllata dal secondo microprogramma è il caricamento nel registro *E* del numero che si trova nel registro *D* se *A* è maggiore o uguale a *B*. In primo luogo si sottrae *B* da *A* e si immagazzina il risultato in *C*. Se il numero in *C* è negativo, ne segue che *A* è minore di *B* e quindi la decisione è di saltare la microistruzione successiva e di riprendere il microprogramma alla microistruzione 4. Se, invece, il risultato della sottrazione è un numero positivo o è zero, *A* è maggiore o uguale a *B*, e il microprogramma procede alla microistruzione successiva, che carica in *E* il numero in *D*. Si aggiunge un quadratino alle microistruzioni per dare il comando di saltare alla microistruzione 4 se *C* è negativo, e corrispondentemente si aggiunge un filo di controllo all'hardware dell'elaboratore.

Si dice che una macroistruzione è «interpretata» dal microcodice per dare i segnali di controllo che manipolano effettivamente il flusso di informazioni nel percorso dei dati. Quindi la successione di microistruzioni nel secondo esempio sopra riportato è l'interpretazione in microcodice della macroistruzione di salto condizionato scritta come «JUMPIF» al passo 19 del programma per la denuncia dei redditi. Il microprogramma affronta il semplice test aritmetico presentato dal programmatore dirigendo l'hardware a controllare il numero nel registro *C* per vedere se è negativo e, in caso di risposta affermativa, lo dirige a eseguire l'opportuna microistruzione di salto condizionato. Per poter operare un controllo del genere l'hardware deve rimandare informazioni dal percorso dei dati al sistema di controllo. L'hardware necessario per determinare se il numero in *C* è negativo è semplicemente un filo partente da *C*, che rappresenta il segno del numero in quel registro: il filo è off (0) quando il segno è positivo (il che significa che *C* è maggiore o uguale a zero) e on (1) quando il segno è negativo (il che significa che *C* è minore di zero).

La sostituzione di semplici strutture ripetitive di memoria al posto di complessi circuiti di controllo cablati porta due vantaggi principali: rende più facile capire e costruire il sistema di controllo e rende più facile modificare il sistema. Un errore in un sistema di controllo microprogrammato può essere corretto semplicemente modificando i contenuti della memoria. Per esempio, supponiamo di decidere di cambiare il microprogramma precedente in modo che carichi in *E* il numero che si trova in *D* non quando *A* è minore di *B*, ma quando *B* è minore di *A*. La prima microistruzione del microprogramma di partenza carica in *C* il risultato di *A–B*. Se però si pone in off il filo 4 e in on il 5, la microistruzione caricherà in *C* il risultato di *B–A*. In ciascun caso la microistruzione successiva determina se il numero in *C* è minore di zero ed effettua il salto se la risposta è affermativa. Quindi si può correggere l'operazione cambiando lo stato di due soli fili di controllo in una microistruzione. Con un sistema di controllo non microprogrammato bisognerebbe modificare la struttura dell'hardware, per poter effettuare la medesima correzione.

Oltre a ridurre il costo di una correzione nel sistema di controllo, la microprogrammazione consente di modificare completamente il sistema di controllo cambiando tutti i contenuti della memoria di microprogramma. Quindi un unico sistema di hardware può svolgere molte funzioni diverse.

I vantaggi del microcodice sono risultati quasi subito evidenti. L'idea di Wilkes, però, precorreva un po' i tempi rispetto alla tecnologia dell'hardware. La microprogrammazione richiede strutture di memoria veloci e poco costose, come non erano disponibili negli anni cinquanta. Inoltre, allora si pensava in genere che un calcolatore progettato in modo da sfruttare questo ingegnoso sistema sarebbe stato più lento di una macchina con sistema di controllo cablato. Di conseguenza la microprogrammazione è rimasta per oltre un decennio poco più che una curiosità intellettuale.

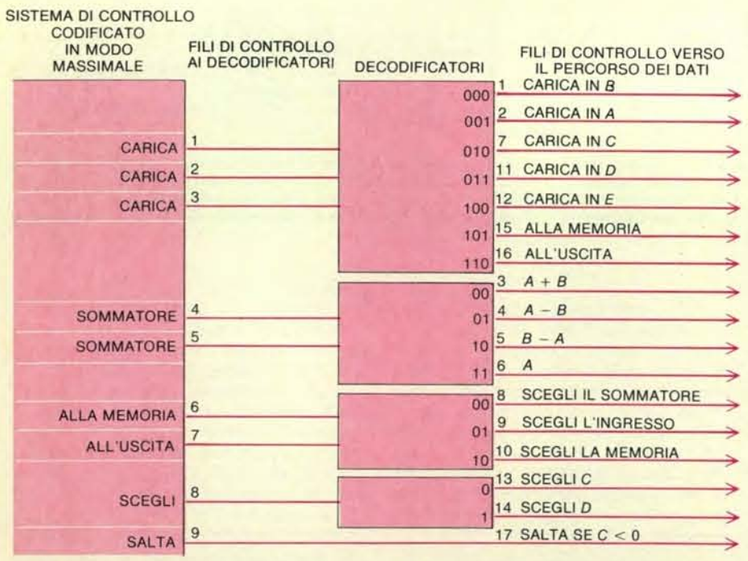
Solo agli inizi degli anni sessanta la microprogrammazione è venuta alla ribalta, in gran parte come esito di una fondamentale decisione commerciale della International Business Machines Corporation (IBM). Verso la fine degli anni cinquanta, la IBM era diventata in effetti una serie di piccole società, ciascuna delle quali commercializzava il proprio tipo di calcolatore. I calcolatori erano diversi non solo per dimensioni, velocità e prezzo, ma anche per l'insieme di macroistruzioni di cui erano dotati: di conseguenza i programmi studiati per un tipo di calcolatore IBM non potevano girare su un altro calcolatore della stessa ditta. Ovviamente sarebbe stato molto più facile vendere calcolatori che funzionassero con un software comune e che differissero solo per dimensioni, velocità e prezzo. Il problema era convincere i dirigenti delle «società» consociate della possibilità, sia per i progettisti dei calcolatori più veloci, sia per i progettisti dei calcolatori più economici, di costruire macchine efficienti basate sullo stesso insieme di macroistruzioni.

I dirigenti delle società consociate furono d'accordo sull'idea innovativa di una famiglia di calcolatori a software compatibile solo perché furono persuasi della realizzabilità di un tale progetto attraverso la microprogrammazione. Se si poteva microprogrammare lo stesso hardware in modo da eseguire insiemi diversi di macroistruzioni (questo fu il ragionamento), allora si doveva poter microprogrammare calcolatori diversi in modo che eseguissero lo stesso insieme di macroistruzioni. I vantaggi commerciali del concetto di famiglia sembrarono tali che la IBM operò un rilevante investimento nello sviluppo di tecnologie di memoria in grado di portare a memorie di microprogramma veloci e poco costose. Infine, nel 1964, la IBM annunciò una famiglia di sette modelli di calcolatori, a software compatibile, che differivano fra loro in velocità per un fattore 300 e in costo per un fattore 100.

Quasi come in un ripensamento, alla IBM si decise che la microprogrammazione avrebbe potuto contribuire anche al compito di rendere compatibili i programmi per i vecchi calcolatori con i nuovi modelli. Di conseguenza diversi modelli della nuova famiglia IBM furono prodotti con due microprogrammi: uno per il nuovo insieme di macroistruzioni e un altro per il vecchio insieme. In questo modo gli utenti potevano far girare i loro vecchi programmi sul nuovo hardware, senza doverli modificare. Il metodo si dimostrò tanto indovinato che per i primi anni sarebbe stato difficile dire se le nuove macchine passavano più tempo a funzionare con il microprogramma per le nuove macroistruzioni o con il microprogramma che emulava le vecchie.

Presto la maggior parte degli altri costruttori di calcolatori seguirono l'esempio della IBM e cominciarono a basarsi sul microcodice per produrre una famiglia di calcolatori a software compatibile formata da macchine molto diverse fra loro per velocità e prezzo. La difficoltà principale incontrata lungo questa strada fu la duplicazione della costosa tecnologia delle memorie inventata dai tecnici della IBM. Tuttavia alla fine il rapido sviluppo della tecnologia microelettronica ha risolto il problema. Per più di un decennio ogni anno l'industria dei semiconduttori è riuscita a raddoppiare la quantità di informazioni immagazzinabile in un unico chip di memoria. Questa tecnologia ha consolidato la posizione della microprogrammazione, come metodo di controllo d'elezione, per tutti gli anni settanta. Le eccezioni sono stati i «supercalcolatori», gli elaboratori più veloci e più costosi: nel loro caso si preferisce ancora un controllo dedicato, per la sua maggiore velocità, anche se è più difficile da costruire.

Quando, circa un decennio fa, apparvero sulla scena i microelaboratori, si pose naturalmente il problema di come progettare il loro sistema di controllo. La storia si è ripetuta: i primi microelaboratori sono stati costruiti con sistemi di controllo non microprogrammati. Con il maturare della tecnologia, però, i vantaggi della microprogrammazione hanno convinto la maggior parte dei progettisti di microcalcolatori ad adottare il microcodice.




**Il numero dei fili di controllo può essere ridotto codificando ulteriormente i segnali di controllo e introducendo opportuni dispositivi di decodificazione. In generale questo metodo comporta il reperimento di due fili che raramente sono attivati insieme per la medesima microistruzione e l'eliminazione di uno di essi, in modo che l'altro possa svolgere una duplice funzione. Questo diagramma schematico dell'hardware raffigura una versione codificata in modo massimale dell'elaboratore raffigurato nelle due illustrazioni a pagina 36. I fili di controllo verso il percorso dei dati sono stati disposti in modo diverso per mettere in evidenza quelli codificati insieme nel nuovo sistema di controllo. L'ulteriore codificazione in genere ha come esito un sistema di controllo più lento di un sistema codificato in modo minimale (un sistema in cui ogni quadrato di microprogramma è messo in corrispondenza con un filo di controllo), perché minore è il numero delle operazioni che possono procedere simultaneamente e perché i segnali di controllo debbono passare attraverso i decodificatori. Inoltre il microprogramma che ne risulta tende ad avere un numero più elevato di microistruzioni. Però, dato che ciascuna microistruzione è composta da un minor numero di cifre binarie, il costo della memoria di microprogramma può ridursi di molto.**



TEST

Sai qual è l'olio giusto per le automobili a gas?

Questa 131 Supermirafiori 1600 cc. è alimentata a GPL. Qual è l'olio piú adatto per questa vettura (e per tutte quelle alimentate a gas)?

83 XOF G1

La risposta esatta alla pagina seguente.

FILI DI CONTROLLO

| MICROISTRUZIONI | 1 CARICA | 2 CARICA | 3 CARICA | 4 SOMMATORE | 5 SOMMATORE | 6 ALLA MEMORIA | 7 ALL'USCITA | 8 SCEGLI | 9 SALTA | COMMENTI |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | CARICA *A* − *B IN C;* VAI A 2 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | SE *C* < 0 VAI A 4 |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | CARICA *D* IN *E;* VAI A 4 |
| 4 | | | | | | | | | | CONTINUA |

**Questo microprogramma, codificato in modo massimale, per il sistema di controllo della pagina 38 viene definito «verticale», perché di solito possiede un numero più elevato di microistruzioni di minore lunghezza il che rende la sua rappresentazione più lunga e più stretta della rappresentazione di un corrispondente microprogramma «orizzontale». In questo caso la codificazione è stata effettuata in modo che non si rendesse necessaria alcuna microistruzione ulteriore.**

Nel frattempo dall'idea della matrice semplice in cui ogni quadratino corrisponde a un singolo filo di controllo si è evoluto un altro tipo di microprogrammazione. I miglioramenti sono venuti soprattutto dal tentativo di ridurre il costo della memoria di microprogramma. Nell'esecuzione di una qualunque microistruzione in genere succede che siano attivi solo pochi componenti dell'hardware; i fili di controllo che portano a tutti gli altri elementi sono off. Ne risulta che un microprogramma tende ad avere isole di 1 sparse qua e là in un mare di 0. Se si potesse raggiungere lo stesso obiettivo con una minore quantità di memoria (cioè con un minor numero di 0), si potrebbe ottenere una memoria meno costosa. La soluzione sta nel fornire la stessa informazione al percorso dei dati codificando ulteriormente i segnali di controllo. In generale l'ulteriore passo di codificazione comporta il reperimento di due fili che raramente sono entrambi attivati per la medesima microistruzione e l'eliminazione di uno dei due, cosicché l'altro può svolgere una duplice funzione. Questa impostazione tende a dare un maggior numero di microistruzioni, ma ciascuna microistruzione ha un minor numero di cifre binarie. Si dice che un insieme di microistruzioni codificato in modo massimale è «verticale», perché i microprogrammi che ne risultano sono in genere lunghi e stretti. Un insieme di microistruzioni codificato in modo minimale viene detto invece «orizzontale» perché i microprogrammi che ne risultano sono in genere corti e larghi. In generale le microistruzioni codificate in modo minimale portano a calcolatori più veloci, perché consentono l'esecuzione simultanea di più operazioni.

Con il crescere dell'importanza della microprogrammazione è aumentato anche l'interesse per metodi che consentano anche agli utenti dei calcolatori di scriversi i propri microcodici. La possibilità di trasformare il calcolatore da un solutore di problemi di carattere generale in una macchina «tagliata» su misura per una ben precisa applicazione ha notevoli vantaggi potenziali, il che ha reso un'idea attraente la microprogrammazione da parte dell'utente. Fin qui i migliori risultati ottenuti da utenti che hanno scritto i propri programmi in microcodice hanno dimostrato un incremento di velocità massimo pari a un fattore 10. Nella pratica il guadagno in velocità dipende dal particolare percorso dei dati del calcolatore microprogrammato e dalla particolare applicazione; in alcuni casi il vantaggio sul piano della velocità è molto inferiore, e a volte può anche capitare che non vi sia alcun miglioramento.

Nella maggior parte dei calcolatori microprogrammati le microistruzioni sono immagazzinate in una sorta di memoria a sola lettura, i cui contenuti sono determinati quando la memoria viene prodotta. Un altro tipo di calcolatore microprogrammato è in grado di modificare i contenuti della sua memoria di microprogramma in pochi millesimi di secondo solamente, mentre il calcolatore è in funzione. Si parla allora di macchine con una memoria di controllo riconfigurabile *(writable control store)*. La chiave di questa impostazione sta nell'avere la disponibilità di molte applicazioni diverse, microcodificate, caricabili solo quando necessario. Dotando il calcolatore di una memoria di controllo riconfigurabile è possibile rimuovere la barriera di hardware che ostacola la microprogrammazione da parte dell'utente.

Nonostante questi tentativi, la microprogrammazione è rimasta così difficile che solo pochi utenti hanno saputo sfruttarne a pieno il potenziale. Per questo alcuni ricercatori hanno affrontato il problema di facilitare la scrittura in microcodice. Sotto molti aspetti la situazione è simile a quella che si venne a creare agli inizi della storia della programmazione. Allora il problema era la difficoltà di disporre migliaia di macroistruzioni in un programma che funzionasse senza falle. Quel problema ha spinto i programmatori a scrivere programmi speciali che aiutassero la stesura dei programmi applicativi. I programmi speciali traducono i linguaggi di programmazione di alto livello nelle macroistruzioni più semplici che un calcolatore è in grado di eseguire. I linguaggi di alto livello come BASIC e Pascal sono più facili da capire e da usare, il che rende meno arduo e faticoso il compito del programmatore, anche se a costo di qualche rinuncia sul piano della velocità del programma.

Negli ultimi anni quanti si occupano di microprogrammazione hanno cercato di creare analoghi programmi di traduzione per il microcodice. Nel nostro caso il problema è più difficile, perché le microistruzioni potenziali sono molto più numerose delle comuni macroistruzioni. Un elaboratore microprogrammato può avere 50 fili di controllo: ogni filo può assumere due stati diversi (on oppure off), e quindi il numero di tutte le possibili microistruzioni sarebbe $2^{50}$, cioè all'incirca $10^{15}$. Inoltre, poiché la velocità del microprogramma determina la velocità del calcolatore, la richiesta di efficienza è molto più pressante. Il risultato è stata l'invenzione di linguaggi di microprogrammazione di alto livello, studiati con l'obiettivo di un equilibrio fra facilità di microprogrammazione ed efficienza. Gli sviluppi recenti sono stati incoraggianti, ma una accettazione generalizzata fra quanti praticano la microprogrammazione sarà possibile solo tra vari anni.

Nel frattempo ricercatori al Thomas J. Watson Research Center della IBM e i miei colleghi e io al dipartimento di ingegneria elettromeccanica e scienza del calcolatore all'Università della California a Berkeley abbiamo seguito una diversa impostazione. Partiamo scartando il tradizionale insieme di macroistruzioni, quindi scriviamo nuovi programmi che traducono i linguaggi di programmazione standard di alto livello direttamente in microcodice. Le macchine basate su questo principio sono chiamate genericamente calcolatori a insieme ridotto di istruzioni *(reduced-instruction-set computer,* o RISC). Per mettere alla prova le nostre idee, a Berkeley abbiamo progettato un nostro microelaboratore, che chiamiamo RISC 1. Inoltre abbiamo semplificato il passo della traduzione codificando le microistruzioni in modo massimale per ottenere un tipo di microprogramma ad andamento estremamente verticale. In effetti, tutti i programmi allora diventano microprogrammi.

Prove effettuate recentemente dal nostro gruppo e da ricercatori della IBM hanno dimostrato che per molte applicazioni un calcolatore a insieme ridotto di istruzioni sarebbe da tre a cinque volte più veloce di un tradizionale calcolatore a insieme complesso di istruzioni. Ricercatori della Stanford University e di vari laboratori industriali hanno dato l'avvio a progetti per lo studio di questa impostazione. Se i dati preliminari saranno confermati, potrà emergere un calcolatore di tipo più semplice, che rappresenterà un inatteso passo avanti nel progressivo sfruttamento del principio della microprogrammazione.

# I fononi

*Come i fotoni sono i «quanti di luce», così i fononi sono i «quanti di suono», entità descrittive astratte alle quali si può ricorrere per determinare le proprietà meccaniche, ottiche e termiche dei solidi*

di Raffaele Guido Della Valle e Pier Francesco Fracassi

In meccanica classica il termine «particella» denota un oggetto fisico dotato di individualità e descrivibile col modello del punto materiale: gli attribuiamo, cioè, una posizione (caratteristica geometrica), una massa (caratteristica statica) e un impulso (caratteristica dinamica). Con l'avvento della meccanica quantistica e il contemporaneo sviluppo dell'indagine sperimentale nel mondo microscopico, il significato del termine «particella» si è progressivamente allontanato dall'idea di «corpuscolo» che originariamente evocava.

Principi quali la dualità onda-particella (Louis De Broglie) e il principio d'indeterminazione (Werner Heisenberg) non solo asseriscono la necessità di un diverso atteggiamento nel nostro modo di osservare e valutare i fenomeni su scala microscopica, ma ci indicano pure le proprietà fondamentali che caratterizzano la «particella quantistica»: un'entità che ci descriva uno dei possibili modi d'essere dell'energia, unitamente alla sua possibilità di trasporto.

Se generalmente oggi il termine particella evoca l'universo dei fenomeni subnucleari (e delle cosiddette «particelle elementari»), la fisica dei solidi ha adottato e sviluppato in modo suo peculiare il concetto di «particella quantistica», intesa quale «eccitazione elementare di un mezzo». La più nota particella di tale tipo è il fotone, l'eccitazione elementare del campo elettromagnetico, responsabile non solo della propagazione della radiazione luminosa, ma pure di tutte le interazioni tra cariche elettriche e momenti magnetici.

Il fonone fu la prima eccitazione elementare in un solido a essere trattata formalmente, con i lavori pionieristici di Max Born e Theodor von Kármán del 1912, nei quali vennero gettate le basi di una teoria dinamica e quantistica dei solidi cristallini (il nome «fonone», o «quanto di suono», venne coniato per analogia, anche etimologica, con «fotone», o «quanto di luce»). In meccanica classica è noto che, se un sistema di corpi in equilibrio stabile viene in qualche modo disturbato, il sistema compie vibrazioni attorno a tali posizioni di equilibrio, e il disturbo si propaga delocalizzandosi sull'intero sistema. Tanto più intense sono le interazioni tra i costituenti del sistema, tanto più veloce sarà tale processo di propagazione (si pensi, per esempio, a un reticolo di palline unite da molle).

Un solido reale, inteso quale sistema di nuclei e di elettroni, obbedisce, nella sua struttura microscopica, alle leggi della meccanica quantistica, ma la sua risposta a disturbi esterni è del tutto analoga: si produce un'eccitazione collettiva delocalizzata, le cui proprietà sono interamente determinate dalla struttura geometrica del solido e dalle forze d'interazione che si manifestano al suo interno.

Allorché tale eccitazione elementare coinvolge principalmente il moto dei nuclei, viene detta «fonone»: l'unità fondamentale di cui sono costituiti tutti i processi elastici e acustici, e buona parte di quelli termici e ottici, che si producono all'interno dei solidi.

La definizione più semplice che si può dare di «solido» è quella di un insieme di atomi o molecole, uniti da forze coesive e localizzati attorno a posizioni d'equilibrio.

I solidi cristallini si distinguono dagli altri stati della materia per la periodicità con cui si organizzano tali posizioni d'equilibrio. Lo spazio all'interno di un cristallo può essere immaginato come suddiviso in tante unità identiche a forma di parallelepipedo: tali unità sono dette «celle unitarie», e l'intera struttura «reticolo cristallino». All'interno di ogni cella unitaria si trova un ugual numero di atomi, ciascuno localizzato attorno a uguali posizioni rispetto alla cella.

Se immaginiamo una ripetizione infinita di tali unità elementari, così da riempirne omogeneamente lo spazio, avremo che le proprietà geometriche e fisiche di ciascuna cella saranno identiche a quelle di una qualunque altra. È tale caratteristica che ci consente di affrontare il problema della descrizione microscopica di tali materiali: la periodicità delle celle unitarie ci garantisce che pure le proprietà dinamiche (e non solo quelle statiche) si ripresentino all'interno del cristallo con analoghe regolarità.

Ovviamente un cristallo reale si discosta da tale modello ideale: le sue dimensioni sono finite, e sono sempre presenti imperfezioni geometriche e impurità chimiche che ne alterano la periodicità. Ma, se le dimensioni del cristallo sono grandi rispetto a quelle della cella unitaria (e il volume tipico di tali celle è inferiore a $10^{-21}$ centimetri cubi) e se le imperfezioni sono sufficientemente «diluite» in un ambiente prevalentemente regolare, sarà possibile descrivere il solido reale e le sue proprietà facendo uso del modello perfetto e infinito; la finitezza dei campioni reali e le loro imperfezioni si riveleranno, nella maggior parte dei casi, inessenziali.

Le proprietà fisiche dei solidi in generale, e dei cristalli in particolare, non sono determinate unicamente dalle loro caratteristiche geometriche. Un ruolo essenziale è svolto dai meccanismi coesivi che regolano la dinamica degli atomi e, quindi, determinano tutte quelle proprietà macroscopiche quali la conduzione termica ed elettrica, il calore specifico, le caratteristice ottiche e magnetiche, e così via. In base alla meccanica quantistica, come il legame chimico nelle molecole, così i meccanismi coesivi nei solidi sono interamente dovuti a interazioni di natura elettromagnetica tra nuclei ed elettroni.

Due sono le difficoltà che si oppongono alla trattazione diretta di tali interazioni: la prima, dovuta all'elevatissimo numero di nuclei e di elettroni in gioco; l'altra, concettualmente più profonda, dovuta al fatto che nuclei ed elettroni sono oggetti quantistici e non classici. Una descrizione che possa essere messa in corrispondenza con un modello classico costituirebbe, se giustificata, una potente guida per lo studio di tali sistemi.

È, dunque, possibile descrivere la dinamica dei nuclei in termini di «oscillazioni attorno all'equilibrio» trascurando i dettagli della dinamica elettronica?

Max Born e Robert Oppenheimer dettero, nel 1927, una risposta positiva a tale domanda, con la loro «ipotesi adiabatica» del moto nucleare: l'idea essenziale è che, tanto nelle molecole quanto nei solidi, i nuclei atomici, di massa notevolmente superiore a quella degli elettroni (un solo protone ha una massa 1018 volte quella dell'elettrone), si muovano molto più lentamente di questi ultimi, cosicché il sistema di elettroni può riadattarsi istante per istante a ogni alterazione delle posizioni nucleari, senza esser costretto a compiere transizioni brusche e improvvise.

Detto in termini più rigorosi, la dinamica dei nuclei coinvolge scale dei tempi più lunghe di quelle coinvolte dalla dinamica elettronica e, se l'ipotesi adiabatica risulta essere un'ipotesi realistica, sarà possibile descrivere la dinamica nucleare come funzione delle coordinate dei soli nuclei.

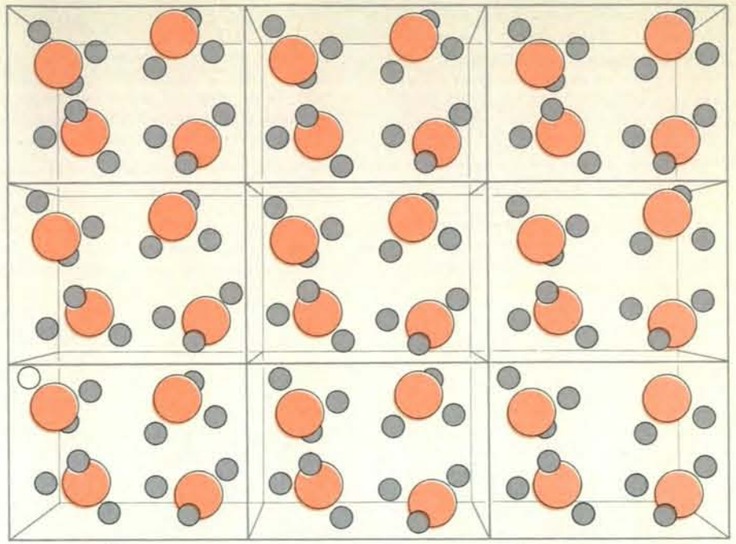
In effetti gli elettroni possono esser classificati grosso modo in due categorie: gli elettroni degli orbitali interni, fortemente influenzati da un unico nucleo e a questo fortemente legati dall'attrazione coulombiana, sì da costituire con esso un sistema essenzialmente rigido detto «nocciolo atomico» (*atomic core*); e gli elettroni, più delocalizzati, degli orbitali esterni, che (come una nuvola di cariche mobilissime) influenzeranno con il loro campo medio il moto dei noccioli atomici.

L'effetto su questi ultimi sarà rappresentabile come un «potenziale efficace» presente tra i vari noccioli atomici e dipendente solo dalle loro coordinate. Il problema originario si è semplificato, in quanto è stato ridotto il numero di coordinate che effettivamente descrivono il sistema (sono state escluse le coordinate degli elettroni), ma è stato introdotto un nuovo tipo di incognita: la forma di questo «potenziale efficace» che guida il moto dei noccioli atomici. Evidentemente è la natura chimica del cristallo a determinarne la configurazione elettronica e, quindi, tale interazione efficace.

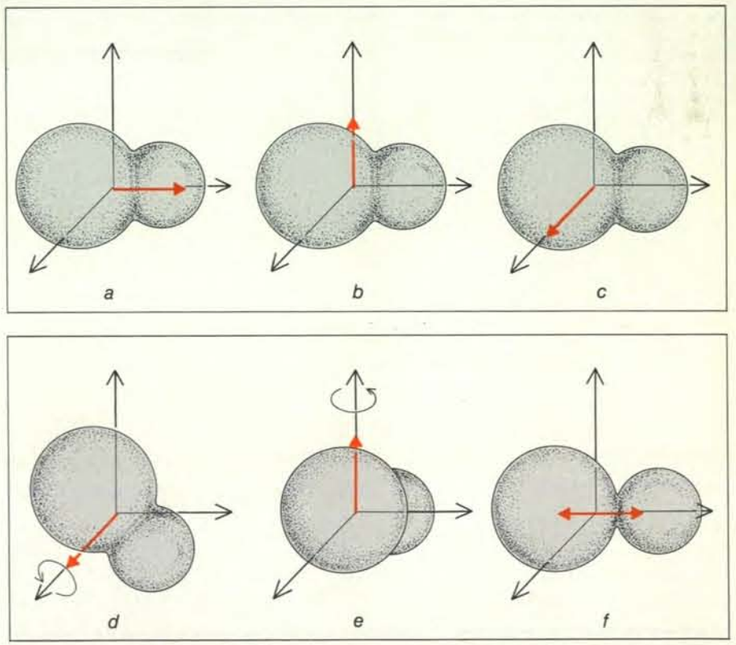
Su tale base, si possono grosso modo distinguere tre tipi principali di solidi cristallini: cristalli atomici o molecolari, in cui tutti gli elettroni sono localizzati attorno a nuclei o gruppi molecolari; cristalli covalenti, dove parte degli elettroni sono delocalizzati sull'intero cristallo e distribuiti su bande di valenza; cristalli metallici, dove gli elettroni delocalizzati sono distribuiti su bande di conduzione e, quindi, presentano un'elevata mobilità.

Una caratteristica fondamentale distingue la classe dei cristalli atomici o molecolari dalle altre due: gli elettroni dei cristalli atomici sono tutti contenuti entro orbitali (atomici o molecolari) saturati, poco influenzati dalla loro interazione col resto del cristallo. Nei solidi covalenti o metallici, invece, i costituenti non presentano orbitali saturi, cosicché, quando vengono uniti per formare il cristallo, gli orbitali elettronici esterni sono violentemente alterati, producendo o legami covalenti distribuiti sull'intero reticolo (quarzo, tormalina, diamante), oppure un mare di elettroni quasi liberi (il «mare di Fermi») tipico dei conduttori elettrici (rame, ferro, argento).

La trattazione dei cristalli costituiti da



**Un cristallo ideale può essere immaginato come suddiviso in tante celle identiche a forma di parallelepipedo, dette «celle unitarie». Ognuna di esse contiene un ugual numero di atomi, ciascuno localizzato attorno a uguali posizioni rispetto alla cella. L'illustrazione mostra la struttura cubica del cristallo di ammoniaca ($NH_3$), dove in ogni cella sono contenute quattro molecole orientate secondo le quattro diagonali del cubo. Un cristallo del genere è detto «molecolare» in quanto le unità $NH_3$ mantengono anche in fase solida la loro individualità: i tre idrogeni (come nel gas) sono uniti da legami covalenti all'atomo di azoto, mentre le interazioni tra molecole diverse, ben più deboli, sono di tipo elettrostatico. Molecole adiacenti sono connesse anche da legami a idrogeno. Questa gerarchia di forze si manifesta nella presenza di due tipi distinti di modi di vibrazione: «interni» (a elevata frequenza, in cui i legami covalenti sono deformati) ed «esterni» (di frequenze tipicamente più basse, che corrispondono a traslazioni e librazioni rigide delle molecole).**

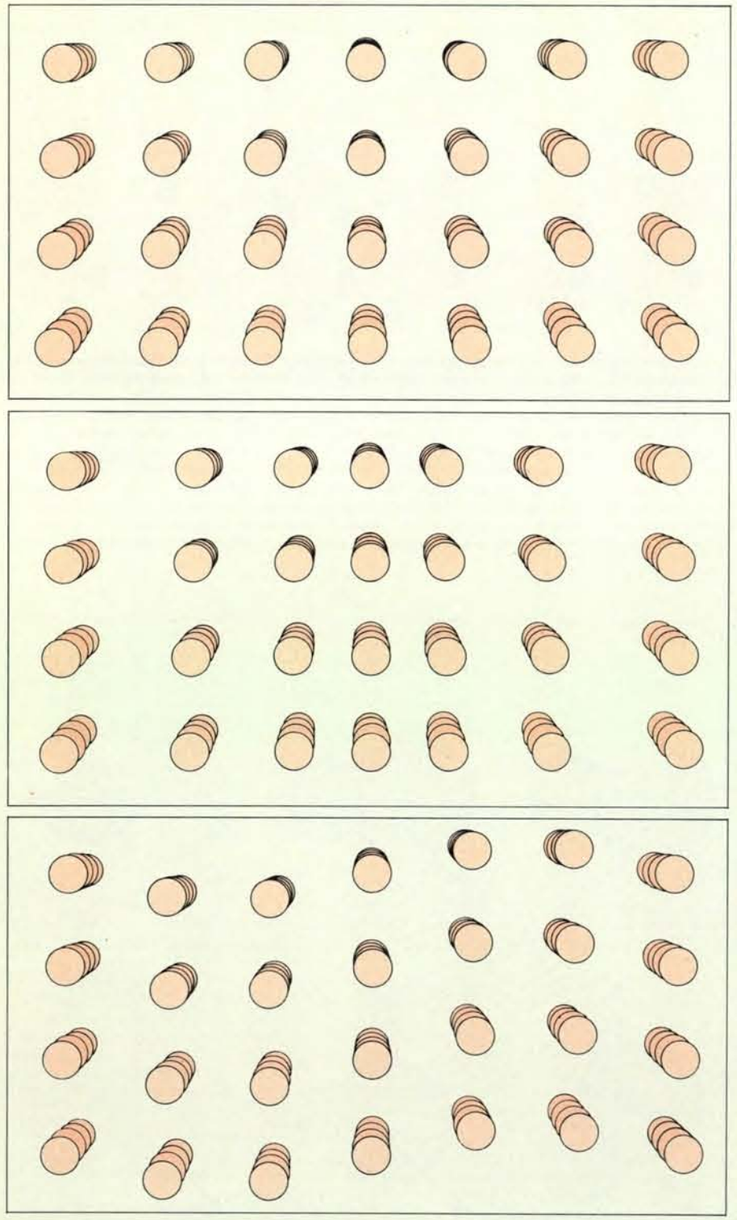



**Una molecola con *N* atomi possiede 3*N* modi indipendenti di moto (gradi di libertà). Una molecola biatomica, come il monossido di carbonio CO, presenta tre traslazioni rigide ortogonali tra loro (*a, b, c*), due rotazioni ortogonali all'asse molecolare (*d, e*), e infine un modo di stiramento lungo l'asse (*f*). Lo stiramento è detto «modo interno» perché, a differenza degli altri, deforma la molecola.**

unità saturate è generalmente più semplice: l'energia d'interazione è prevalentemente di natura elettrostatica e i suoi aspetti non classici possono esser calcolati con i metodi della meccanica quantistica, cosicché la determinazione di potenziali efficaci realistici e, entro certi limiti, attendibili è facilitata.

Se sono presenti unità ioniche (come accade nei cristalli di cloruro di potassio, $K^+Cl^-$, o di nitrito di sodio, $Na^+NO^-_2$) l'interazione prevalente è a lungo raggio, di tipo coulombiano. Nel caso di molecole elettricamente neutre, saranno le disomogeneità nella distribuzione interna delle cariche a determinare forze elettrostatiche a medio raggio (acqua, ammoniaca, benzene).

Per quanto riguarda, infine, i solidi di gas nobile, la stabilità chimica delle unità atomiche consente unicamente forze debolissime a corto raggio (dette «di van der Waals») originate dalle fluttuazioni delle nuvole elettroniche e da fenomeni, tipicamente quantistici, di scambio elettronico.



**Quando la periodicità di un cristallo all'equilibrio *(in alto)* viene alterata da una deformazione, si producono al suo interno forze di richiamo che instaurano vibrazioni collettive di tipo ondulatorio degli atomi. In un cristallo monoatomico si osservano due soli tipi di onde: onde longitudinali di compressione e rarefazione *(al centro)*, nelle quali il movimento oscillatorio degli atomi è parallelo alla direzione di propagazione dell'onda, e onde trasversali *(in basso)*, tipiche dei solidi, dove invece il moto degli atomi è ortogonale alla direzione di propagazione.**

A questo punto introduciamo un modello teorico che permette di risolvere con relativa facilità il problema del moto dei noccioli atomici. Consideriamo una molecola biatomica: questa può essere descritta come composta da due noccioli atomici uniti da una forza dipendente dalla loro interdistanza e che rappresenta il legame covalente. Tale molecola presenta sei gradi di libertà. Può, cioè, traslare rigidamente in tre direzioni ortogonali *(x, y, z)*, ruotare attorno ai due assi ortogonali all'asse molecolare (ossia all'asse passante per i due nuclei) e vibrare parallelamente a quest'ultimo.

L'esistenza di una distanza d'equilibrio tra i due noccioli atomici significa che, se questi si avvicinano, si produrrà tra loro una forza repulsiva (e, viceversa, attrattiva se si allontanano) con il risultato che nella molecola potranno instaurarsi vibrazioni caratterizzate dalla forza coesiva del legame covalente.

Le leggi secondo cui tale forza varia con la distanza sono, in genere, molto complesse; ma, se l'ampiezza delle oscillazioni è piccola, tale forza risulta essere semplicemente proporzionale alla variazione $d$ dell'interdistanza, ossia $F = -kd$, dove la costante $k$ (che esprime l'intensità del legame covalente) dipende dalla natura dei due atomi coinvolti e il segno meno è per esprimere che con $d$ positivo (allontamento degli atomi) $F$ è negativa (ossia tendente a riavvicinarli), e viceversa. Un sistema meccanico che si comporti in tale modo è detto «oscillatore armonico», e una molecola biatomica, allorché compie piccole oscillazioni, è con buona approssimazione un oscillatore armonico.

Se i due noccioli atomici obbedissero alle leggi della meccanica classica (come due palline unite da una molla), il moto periodico avrebbe frequenza $\nu = \sqrt{k}$ e sarebbe realizzabile qualunque ampiezza (e, quindi, qualunque energia) d'oscillazione. La natura quantistica degli atomi, invece, impone delle limitazioni, e precisamente sono attuabili solo quegli stati per cui l'energia sia un multiplo intero di $h\nu$, dove $h = 6{,}62 \times 10^{-27}$ erg al secondo è la costante di Planck. Quando la molecola è in uno di questi stati stazionari, gli elettroni seguono adiabaticamente (cioè senza compiere transizioni brusche) il moto dei nuclei.

La molecola può passare da uno stato stazionario a un altro con un salto quantico, scambiando con l'esterno frazioni d'energia pari a multipli interi di $h\nu$; per esempio, in un gas ciò può accadere come conseguenza di un urto con un'altra molecola.

Allorché a tale moto vibratorio si associa uno spostamento in controfase di cariche elettriche di segno opposto (come accade, per esempio, nella molecola di monossido di carbonio, CO, dove, per la maggior elettronegatività dell'ossigeno rispetto al carbonio, esiste un addensamento di carica negativa attorno al nucleo d'ossigeno e positiva attorno a quello di carbonio), la vibrazione può interagire con il campo della radiazione elettromagnetica: la molecola può, cioè, irradiare emettendo un fotone di frequenza $\nu$ ed energia $h\nu$, oppure può aumentare il suo stato vibrazionale assorbendo un fotone di ugual frequenza ed energia.

Del tutto analoga è la situazione nel caso di molecole poliatomiche: per piccole oscillazioni è valida l'approssimazione armonica (ossia le forze di richiamo che si producono a seguito di deformazioni sono proporzionali alle deformazioni stesse); e, se a tali deformazioni si associano spostamenti di cariche elettriche, le vibrazioni potranno «accoppiarsi» al campo elettromagnetico della radiazione: la molecola potrà assorbire o emettere fotoni effettuando una transizione tra due diversi stati vibrazionali.

Il maggior numero di nuclei determina un maggior numero di «modi di vibrazione» possibili, caratterizzati da frequenze distinte associate a modi distinti di deformazione della molecola: se la molecola contiene $N$ atomi, saranno possibili $3N$ modi di moto.

Caratteristica di tali vibrazioni armoniche è la loro indipendenza reciproca: lo stato di vibrazione a frequenza $\nu'$, ad esempio, non influenza il moto a frequenza $\nu''$, e può coesistere con esso.

Allorché tale metodologia è applicata a un sistema cristallino, viene affrontato uno dei temi principali della fisica dei solidi: la «dinamica reticolare».

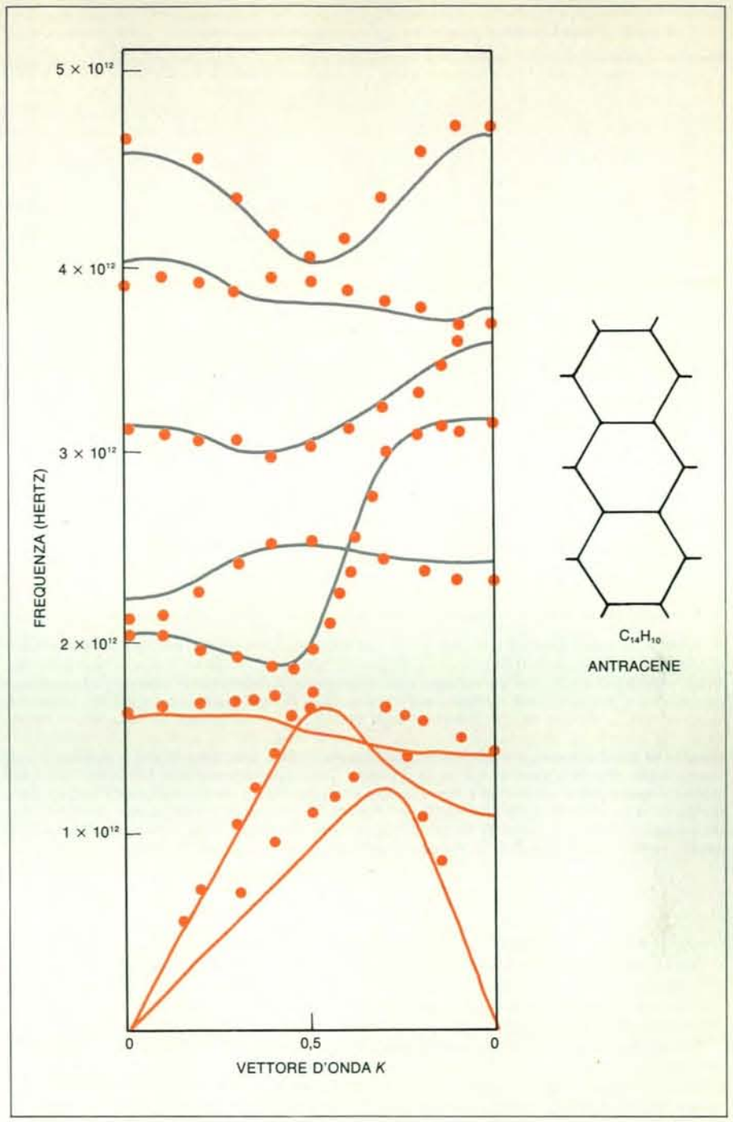
Anticipata dai lavori sulla teoria quantistica dei calori specifici di Albert Einstein (1907) e di Peter Debye (1912), la dinamica reticolare venne affrontata per la prima volta da Max Born e Theodor von Kármán nel 1912, ma il suo vero contenuto fisico non fu pienamente compreso per più di trent'anni, e solo nel 1954 la teoria raggiunse la sua prima maturità con il trattato di Max Born e Kun Huang *Dynamical Theory of Crystal Lattices*.

Se supponiamo che le forze agenti tra i noccioli atomici di un cristallo siano di natura armonica, anche in tale struttura si realizzeranno movimenti oscillatori indipendenti l'uno dall'altro.

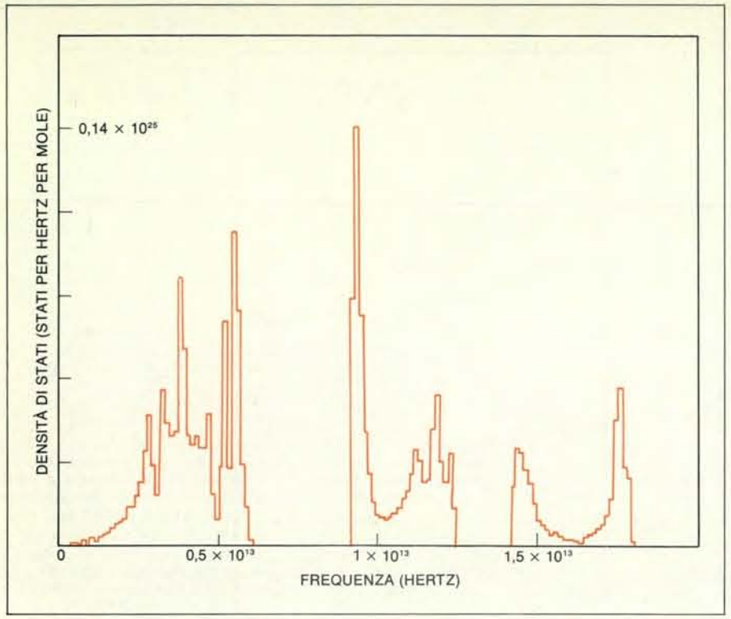
Se il reticolo di un cristallo macroscopico è costituito da $N$ atomi, i modi distinti di moto saranno, come per le molecole, $3N$; unica differenza è che, adesso, $N$ è tipicamente un numero dell'ordine di $10^{23}$ (ossia del numero di Avogadro): il calcolo delle frequenze e dei modi parrebbe impraticabile, dovendo esser risolte $10^{23}$ equazioni di moto!

Fortunatamente la struttura periodica del cristallo permette semplificazioni decisive: difatti ciascuna cella, identica a tutte le altre, presenta gli stessi movimenti possibili (alle medesime frequenze) di una qualunque altra cella. Ciò farà sì che lo stato dinamico di una cella si propaghi, identico, alle celle vicine: le oscillazioni del cristallo nel suo insieme si potranno descrivere in termini di onde elastiche. Sarà sufficiente risolvere le equazioni di moto per un'unica cella, e il moto delle altre potrà essere dedotto con semplici considerazioni geometriche.

Per comprendere, almeno qualitativamente, i moti oscillatori che hanno luogo in un solido cristallino, immaginiamo che cosa accade a una molecola (ad esempio il monossido di carbonio, CO) quando en-




**La dinamica reticolare di Max Born ebbe una convincente conferma sperimentale nel 1955, quando le prime esperienze di spettroscopia a neutroni, condotte da B. N. Brockhouse a Chalk River, in Canada, mostrarono l'esistenza dei fononi e la dipendenza della loro frequenza dal vettore d'onda. Nel grafico sono riportate le frequenze, in funzione del vettore d'onda, dei modi normali del cristallo molecolare di antracene. I valori sperimentali *(pallini)* sono stati misurati con spettroscopia a neutroni da I. Natkaniec a Dubna, in Unione Sovietica, mentre le curve teoriche sono state calcolate dagli autori. Le tre curve inferiori *(in colore)* corrispondono ai modi acustici (traslazioni rigide delle molecole), mentre le altre corrispondono ai modi ottici. Questi ultimi sono complesse combinazioni di librazioni, traslazioni e deformazioni delle molecole.**




**Il numero di modi normali presenti in un cristallo macroscopico è dell'ordine del numero d'Avogadro. Come si distribuiscono in frequenze tali modi normali? La loro distribuzione, detta «densità di stati», non obbedisce a una legge generale, bensì è determinata dalla struttura geometrica e dai potenziali d'interazione caratteristici di ogni particolare cristallo. L'illustrazione mostra la densità di stati calcolata dagli autori per i modi esterni di librazione e traslazione del cristallo di ammoniaca: in ascissa vi sono le frequenze, in ordinata il corrispondente numero di modi normali. I modi interni, non rappresentati, sono tutti situati a frequenze nettamente più alte (maggiori di 3,6 × $10^{13}$ hertz). Tale separazione tra le frequenze dei modi esterni e quelle dei modi interni è conseguenza del fatto che i legami covalenti all'interno delle molecole sono estremamente più rigidi dei legami, di natura prevalentemente elettrostatica, tra molecole diverse. La densità di stati di un cristallo determina tutte le proprietà termodinamiche, come i calori specifici, le energie di sublimazione, le variazioni d'entropia, e così via.**

tra a far parte di un reticolo. Come si è detto, la molecola biatomica libera nel gas presenta sei tipi indipendenti di movimento: tre traslazioni, due rotazioni e una vibrazione interna di stiramento.

Nel gas le velocità sono in genere così elevate, e la densità così bassa, che due molecole possono influenzare reciprocamente i loro moti solo nell'eventualità di urti. In un reticolo cristallino (e, in genere, in ogni fase condensata) le ridotte velocità e la contiguità di più molecole rendono le mutue interazioni (nel nostro caso di tipo prevalentemente elettrostatico) capaci di influenzare continuamente i moti molecolari; gli originali moti di traslazione saranno limitati a oscillazioni rigide dell'intera molecola, producendo (nel loro propagarsi) tre onde di compressione e rarefazione (una longitudinale, come nei fluidi, e due trasversali tipiche dei solidi); le rotazioni degenereranno in moti di «beccheggio» (librazioni) delle molecole attorno a orientazioni di equilibrio; lo stiramento, infine, conserverà più degli altri modi le proprie caratteristiche, con la condizione, però, di non poter più interessare un'unica molecola, ma di comunicarsi all'intero sistema. Quelli che erano sei gradi di libertà nella molecola libera di CO sono divenuti sei tipi distinti di onde elastiche nel rispettivo cristallo.

L'ipotesi armonica, che garantisce l'indipendenza di oscillazioni distinte, ci permette di trattare tali onde come non interagenti e, quindi, di esaminarle individualmente.

La presenza di un'onda elastica denota l'esistenza di deformazioni del reticolo periodiche nel tempo e nello spazio. Ciascuna onda è caratterizzata da una frequenza $\nu$ e da un vettore d'onda $k$, cioè un vettore avente per direzione quella di propagazione e per modulo il reciproco della lunghezza d'onda. Ciascuna di queste onde elastiche indipendenti, ossia ciascuno di questi movimenti collettivi interessanti l'intero cristallo, è detto «modo normale» di vibrazione.

A causa della disomogeneità e anisotropia di ogni reticolo cristallino, la frequenza di tali modi normali dipenderà non solo dal tipo di deformazione periodica considerata, ma anche dalla sua lunghezza d'onda e dalla direzione di propagazione, ossia dal vettore d'onda $k$. Il numero possibile di modi dipende dalla struttura della cella unitaria e in generale, se $m$ è il numero di atomi per cella, vi saranno $3m$ modi normali distinti per ogni vettore d'onda (per esempio, l'ammoniaca, $NH_3$, che cristallizza in un reticolo cubico con quattro molecole per cella, presenta 48 modi distinti), ma comuni a tutte le strutture sono i tre modi di compressione e rarefazione detti «acustici» e caratterizzati da traslazioni rigide in fase di tutti gli atomi appartenenti alla medesima cella unitaria.

I modi acustici a grande lunghezza d'onda sono sempre a bassa frequenza e determinano le proprietà acustiche e quelle elastiche quasi-statiche (come la comprimibilità) del materiale. Gli altri modi (librazioni, vibrazioni interne molecolari e traslazioni in controfase) sono detti «ottici».

Il significato di tale denominazione appare chiaro se si considera che per tali modi, a differenza dei modi acustici, è possibile l'instaurarsi di vibrazioni in controfase di cariche elettriche di segno opposto, con la conseguente capacità d'interazione con il campo elettromagnetico della radiazione.

Finora abbiamo trascurato la natura quantistica del cristallo. In realtà per ogni modo normale non sono permesse tutte le energie, come accadrebbe in un sistema classico, bensì per ogni frequenza $\nu$ si potranno realizzare solo le oscillazioni per cui l'energia è un multiplo intero di $h\nu$. Abbiamo anche visto come, per ciascun modo, la frequenza cambi al variare del vettore d'onda $k$ (ossia al variare della lunghezza d'onda e della direzione di propagazione).

Lo stato dinamico complessivo del cristallo sarà determinato completamente allorché, per ogni modo di frequenza $\nu$, sarà specificato (con un numero intero $n$) il livello a cui esso è eccitato. Tale numero $n$ è detto «numero d'occupazione», e il rispettivo modo avrà un'energia pari a $n$ volte $h\nu$.

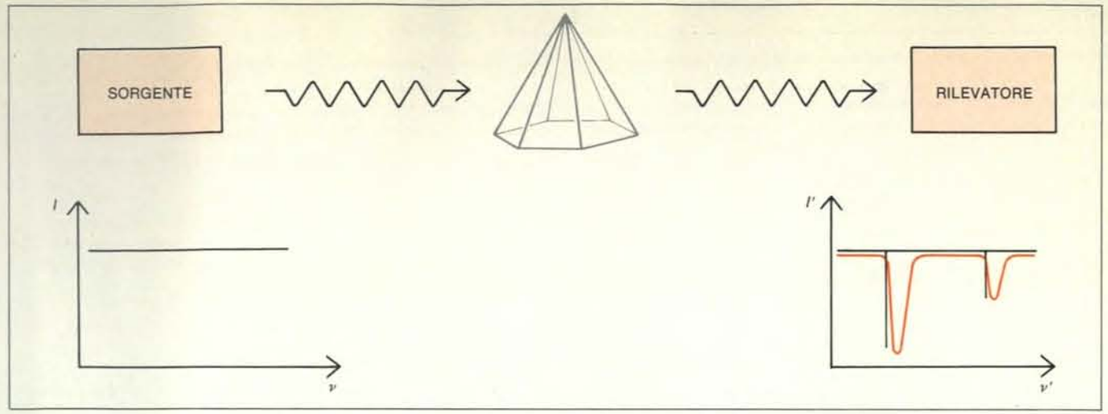
Ciascuna di queste unità $h\nu$ d'eccitazione vibrazionale del reticolo cristallino è detta «fonone» e quindi lo stato dinamico di un cristallo si esprime fornendo il numero d'occupazione (ossia, il numero di fononi presenti) per ciascun modo del reticolo.

Come si vede, all'origine del concetto di fonone vi è essenzialmente un'astrazione matematica, giustificata unicamente dalla natura delle soluzioni trovate al problema della dinamica reticolare.

L'entità che è stata chiamata «fonone» (e lo stesso si potrebbe ripetere per altre eccitazioni elementari, quali sono i fotoni per il campo elettromagnetico, o gli eccitoni, i plasmoni, i magnoni nei solidi) è in primo luogo un «elemento descrittivo» suggerito dal modello astratto con cui si rappresenta un mezzo fisico reale e con cui si cerca di descrivere le proprietà osservabili di quest'ultimo. A ogni fonone sono associate quelle capacità di trasporto d'energia e di impulso, che si richiedono a una «particella quantistica»; possiamo, cioè, descrivere ciascun fonone come una «particella» dotata di energia $h\nu$ e impulso $hk$.

L'uso di tali particelle come entità






**Una misura spettroscopica d'assorbimento viene realizzata facendo attraversare il cristallo da un fascio di fotoni emesso da una sorgente, mentre uno strumento di rilevamento osserva la radiazione uscente. Le proprietà della sorgente sono note, cosicché conosciamo l'intensità *I* della radiazione per ogni frequenza $\nu$ incidente sul cristallo. Il cristallo, interagendo con la radiazione, altera tale distribuzione iniziale d'intensità, producendo una distribuzione *I'* che viene misurata dal rilevatore. Il rapporto tra le due distribuzioni (lo «spettro») contiene, ovviamente, informazioni circa le proprietà del cristallo. Tali informazioni possono essere decodificate solo disponendo di un modello teorico che descriva il solido e le sue modalità d'interazione con la radiazione. Gli spettri normalmente osservati presentano bande o picchi di assorbimento di larghezze finite *(in colore)*; il modello armonico, invece, prevede righe infinitamente sottili *(in nero)* in prossimità dei massimi dei picchi sperimentali. Solo modelli che tengano conto dell'anarmonicità (e quindi delle interazioni tra fononi) possono spiegare le larghezze finite e la forma delle bande osservate. In tal senso la teoria anarmonica dei solidi cristallini è qualitativamente superiore a quella puramente armonica.**

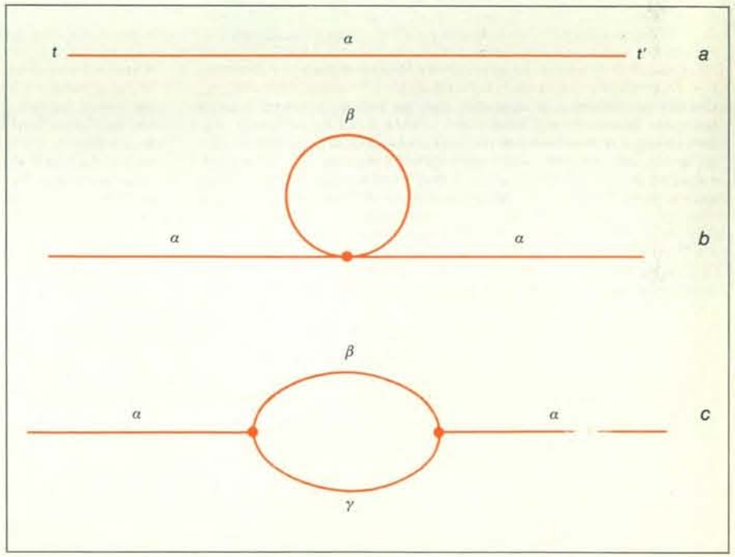
dotate, in qualche modo, di una realtà oggettiva viene reso legittimo dal fatto che, nel comportamento collettivo di tali eccitazioni elementari, si ritrovano quei fenomeni che si osservano sperimentalmente nei solidi.

Abbiamo visto come sia possibile, per il campo elettromagnetico, agire sulle componenti atomiche di un cristallo eccitandovi i modi ottici di vibrazione, ossia quei movimenti che presentano vibrazioni nella distribuzione di cariche elettriche di segno opposto.

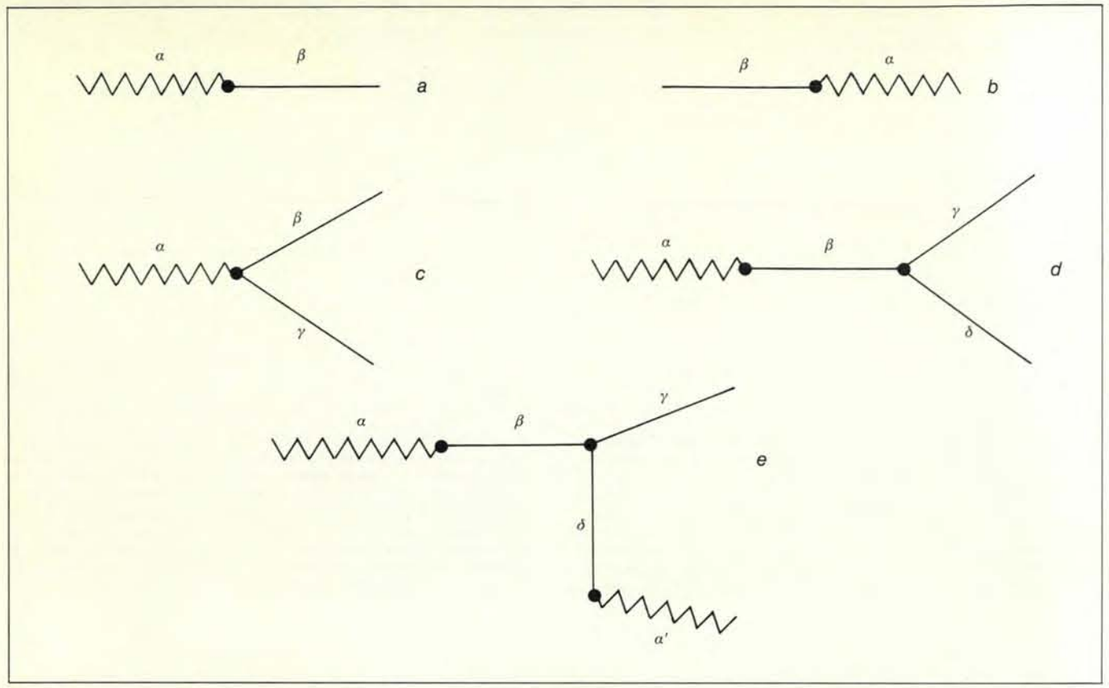
Se si vuole descrivere tutto ciò con il formalismo delle eccitazioni elementari, si possono trattare i vari fenomeni fisici come sovrapposizioni di processi più semplici, la cui rappresentazione è intuitivamente più immediata. Difatti, il campo elettromagnetico della radiazione può essere descritto in termini di fotoni, che altro non sono se non le «particelle quantistiche» associate alle eccitazioni elementari dei modi normali del campo: in completa analogia con un fonone, un fotone di frequenza $\nu$ e vettore d'onda $k$ trasporta un'energia $h\nu$ e un impulso $hk$.

Ne consegue che le complesse interazioni tra radiazione e cristallo possono essere riformulate in termini di interazioni tra fotoni e fononi, ossia possono essere analizzate in processi elementari nei quali alcune «particelle» sono create e altre distrutte, conservando l'energia e l'impulso complessivi. L'assorbimento di radiazione, nella sua descrizione microscopica, risulta un processo in cui un fotone viene distrutto e viene creata nel cristallo una sua eccitazione elementare.

Tre sono le condizioni da soddisfare affinché tale processo abbia luogo: conservazione dell'energia; conservazione dell'impulso; possibilità d'accoppiamen-




**I diagrammi di Feynman sono un potente ausilio per la fisica teorica delle particelle elementari e dello stato solido. Benché la loro struttura grafica evochi l'idea di traiettorie e collisioni, ciascun diagramma è, in realtà, la rappresentazione di un'espressione matematica ben definita, esprimente la probabilità che a un determinato stato iniziale (rappresentato dalle linee che «entrano» nel diagramma) consegua un particolare altro stato finale (rappresentato dalle linee che «escono» dal diagramma): ciascuna linea è associata a un'eccitazione elementare del sistema, e ciascun vertice (punto d'incontro di più linee) rappresenta le loro interazioni possibili. Regole ben precise permettono, poi, di dedurre da ogni configurazione un'appropriata espressione algebrica. La propagazione «libera» di un fonone (cioè in assenza di interazioni con altri fononi) è rappresentata dal diagramma *a*, cui corrisponde la descrizione (espressa in termini di correlazioni di probabilità) di come il fonone $\alpha$ si propaghi dall'istante *t* all'istante *t'*. Se si introduce l'interazione tra fotoni, alla probabilità della propagazione libera va aggiunta quella dovuta all'eventualità di processi di tipo *b* (diffusione elastica del fonone $\alpha$ dovuta al fonone $\beta$) e di tipo *c* (dove la propagazione $\alpha$ è «interrotta» dal prodursi di una coppia «virtuale» $\beta$ e $\gamma$) e quindi da processi via via più improbabili descritti da diagrammi sempre più complessi. Processi del tipo *b* alterano la frequenza del fonone $\alpha$, mentre processi del tipo *c* rendono finita la sua vita media. L'efficacia di tali processi è determinata dai potenziali d'interazione all'interno del cristallo e dalla sua temperatura; calcoli numerici completi di tali effetti si sono resi possibili solo recentemente e richiedono l'uso di calcolatori molto veloci.**




**I diagrammi di Feynman possono essere impiegati pure per descrivere le interazioni tra fotoni *(linee ondulate)* e fononi *(linee diritte)*. Allorché un fotone $\alpha$ è risonante con un fonone $\beta$ (ovvero ha la medesima frequenza e il medesimo vettore d'onda), potranno prodursi processi d'assorbimento *(a)* o processi inversi di emissione *(b)*. Un fotone non risonante può essere ugualmente assorbito dal cristallo tramite processi multifononici: il diagramma *c* rappresenta il processo in cui il fotone $\alpha$ crea direttamente due fononi $\beta$ e $\gamma$ di energia e impulso complessivi uguali a quelli del fotone distrutto. La probabilità che un tale evento si verifichi è spesso esigua; molto più efficace è il processo indiretto *(d)* in cui il fotone crea un fonone virtuale $\beta$ a energia diversa dalla propria: simili violazioni della conservazione dell'energia sono consentite nei sistemi quantistici, ma solo per tempi estremamente brevi; il fonone virtuale decade immediatamente in due fononi $\gamma$ e $\delta$ con energia e impulso complessivi uguali a quelli del fotone iniziale. Ciascuna di tali transizioni, presa singolarmente, non ha elevata probabilità, ma, poiché il numero di transizioni possibili con uguale stato finale è elevatissimo, il processo complessivo è molto efficace. Il diagramma *e*, infine, rappresenta la diffusione anelastica *(scattering* Raman) della radiazione, in cui la frequenza del fotone incidente $\alpha$ è diversa da quella del fotone riemesso $\alpha'$. L'energia mancante è dovuta alla contemporanea eccitazione di un fonone $\delta$ tramite due fononi virtuali $\beta$ e $\gamma$. Tale fenomeno è sfruttato nella cosiddetta «spettroscopia Raman» per la misura sperimentale delle frequenze dei fononi (in alternativa a misure di spettri d'assorbimento), utilizzando sorgenti di radiazione laser per il fotone incidente $\alpha$.**

to. La prima condizione impone che la frequenza dell'eccitazione elementare prodotta sia la stessa del fotone distrutto; le frequenze tipiche dei fononi vanno da $10^{10}$ a $10^{14}$ hertz, quindi l'interazione d'assorbimento è limitata ai fotoni dell'infrarosso e dell'infrarosso lontano. Poiché la velocità di propagazione dei fotoni è quella della luce ($3 \times 10^{10}$ centimetri al secondo), le loro lunghezze d'onda saranno molto superiori alle dimensioni delle celle unitarie, ovvero i vettori d'onda saranno molto piccoli. L'impulso trasportato da un'eccitazione di vettore d'onda $k$ è $hk$ e, quindi, per la conservazione dell'impulso, solo quei pochi fononi con piccolo vettore d'onda potranno interagire con i fotoni.

Come avviene l'interazione nei casi permessi? Diamone prima una descrizione ondulatoria. Come abbiamo visto, i modi ottici del cristallo sono vibrazioni caratterizzate da campi elettrici e magnetici oscillanti prodotti dalle cariche stesse del reticolo. Allorché giunge un'onda elettromagnetica, questa perturberà il moto delle cariche e sarà da esse perturbata a sua volta, cosicché nel cristallo si propagherà un disturbo la cui energia è parte in forma elastica (come un fonone) e parte in forma elettromagnetica (come un fotone). Se la frequenza originaria della radiazione è lontana da ogni possibile frequenza reticolare, la componente «fononica» è assente (il cristallo è trasparente). Viceversa, se la frequenza è esattamente uguale a una frequenza ottica del reticolo (risonanza), la componente «fotonica» sarà nulla (il cristallo assorbe radiazione).

Nei casi intermedi avremo dei modi «ibridi» elastici ed elettromagnetici al tempo stesso. Poiché pure ora valgono le regole di quantizzazione, tali modi ibridi possono trasportare solo quantità discrete di energia del valore di $h\nu$, se $\nu$ è la loro frequenza. Tali modi accoppiati non risonanti fonone-fotone sono detti «polaritoni»; tanto più è grande la loro componente fotonica, tanto più la loro propagazione è di tipo ottico; se prevale la natura fononica, la loro propagazione è di tipo elastico e, al limite della perfetta risonanza, tutta l'energia elettromagnetica del fotone è convertita in energia elastica del reticolo: si ha un processo di assorbimento.

Esprimendo tutto ciò in termini di «particelle quantistiche», diciamo che i fotoni (ovvero il campo elettromagnetico di radiazione esterno al cristallo) possono decadere in eccitazioni elementari del solido: a seconda delle frequenze in giuoco, queste saranno fononi (eccitazioni puramente elastiche), polaritoni (eccitazioni ibride elastico-elettromagnetiche che realizzano il campo di radiazione all'interno del cristallo), ovvero, per frequenze nel visibile e oltre, eccitazioni elettroniche collettive come gli eccitoni e i plasmoni.

All'inizio degli anni cinquanta si svilupparono, a opera principalmente di Richard P. Feynman e Freeman J. Dyson, i metodi della meccanica quantistica dei campi, che permisero la prima trattazione teorica quantitativa delle particelle elementari. Negli stessi anni, per merito principalmente di A. A. Abrikosov, L. P. Gorkov e della scuola russa, tali metodi vennero estesi alla meccanica statistica, rendendo possibile lo studio delle interazioni tra le eccitazioni elementari nei solidi. Ciò produsse, tra l'altro, una nuova esplosione d'interesse nei confronti della dinamica reticolare, e vennero effettuati, da Alexei Maradudin, E. R. Cowley e collaboratori, i primi calcoli che tenessero conto della cosiddetta «interazione anarmonica» tra fononi.

L'indipendenza dei modi normali di un cristallo si traduce nell'assenza di interazioni tra i vari fononi: allorché un fonone viene creato eccitando un modo normale, questo si manterrebbe indefinitamente qualunque sia lo stato dinamico del resto del cristallo. Tale indipendenza deriva unicamente da una semplificazione formale che fu fatta all'inizio della trattazione: l'ipotesi armonica. L'ipotesi, cioè, che per piccole oscillazioni le forze di richiamo siano proporzionali alle deformazioni del reticolo.

Benché tale ipotesi sia sufficientemente accurata da fornire stime ragionevoli delle frequenze dei vari modi normali, la sua validità non è però assoluta; i termini «anarmonici» dell'interazione tra atomi e molecole (ossia tutte quelle componenti degli sforzi che non sono linearmente proporzionali alle deformazioni) producono correzioni alla dinamica armonica, che si traducono in interazioni tra fononi. Si avranno, così, eventi in cui un fonone è distrutto e se ne creano due con energia e impulso complessivi uguali a quelli del fonone originario, processi inversi (in cui, cioè, due fononi sono distrutti e uno creato) e, poi, processi via via più improbabili che coinvolgono un numero sempre maggiore di fononi.

La presenza di tali interazioni perturba l'evoluzione di ogni eccitazione elementare, che è così influenzata dal continuo scambio di impulso e di energia tra i vari modi normali. Due sono gli effetti immediati: l'alterazione della frequenza di ogni fonone, e l'introduzione di una possibilità di decadimento che ne rende finita la vita media. E così, mentre il modello armonico prevedeva assorbimento della radiazione unicamente a frequenze esattamente definite (con conseguenti spettri a righe infinitamente strette), le vite medie finite si rivelano, sia sperimentalmente che teoricamente, con righe d'assorbimento di larghezza finita.

Solo negli ultimi anni una conoscenza più approfondita delle forze interatomiche e intermolecolari, unita alla disponibilità di calcolatori molto veloci, ha reso possibile il calcolo teorico di tali effetti anche in cristalli complessi quali sono quelli molecolari.

L'indagine teorica su tali materiali si sta rapidamente sviluppando, attraverso una vasta collaborazione su scala internazionale, che vede impegnati, tra gli altri, Maradudin, a Irvine, California; Michael Klein, al National Research Council di Ottawa; Stuart Pawley, a Edimburgo; Stuart Walmsley, all'University College di Londra; Salvatore Califano, a Firenze; Tadeusz Luty, a Cracovia; e i loro collaboratori.

Sin qui non ci siamo chiesti quanti fononi popolino «normalmente» un cristallo. In effetti esiste una precisa relazione tra il numero medio di fononi presenti in un solido in equilibrio termico e la sua temperatura $T$. Tale relazione è stabilita, come per i fotoni, dalle regole della statistica di Bose-Einstein, la quale asserisce che il numero medio $n$ di fononi di frequenza è dato dalla relazione di Planck $n = (\exp(h\nu/kT) - 1)^{-1}$, dove $k$ è la costante di Boltzmann pari a $1{,}38 \times 10^{-16}$ erg/kelvin.

Tale popolazione termica media dei fononi è detta «distribuzione di equilibrio». È evidente, quindi, il ruolo determinante che ha la temperatura del cristallo per tutti quei processi che coinvolgono interazioni con i fononi; in particolare, temperature più elevate, ovvero popolazioni medie più grandi, aumentano la probabilità di decadimento delle eccitazioni elementari, riducendone la vita media.

Allorché un disturbo esterno (per esempio, un fascio di fotoni) altera lievemente tale distribuzione d'equilibrio, creando o distruggendo alcuni fononi, il sistema tenderà a ritornare allo stato d'equilibrio termico attraverso i processi d'interazione fonone-fonone. Lo scambio d'energia tra il sistema e l'agente esterno verrà determinato da tale evoluzione verso l'equilibrio, che sarà tanto più efficiente e veloce, quanto più intense sono le interazioni tra i fononi e quanto più consistente è la popolazione d'equilibrio preesistente.

L'osservazione sperimentale di un fonone (o, più in generale, di una qualsiasi eccitazione elementare in un solido) consiste, quindi, nell'alterazione della popolazione d'equilibrio e nella conseguente risposta del sistema.

L'adozione di un modello corpuscolare per la descrizione di tali processi è stata originata dall'evidenza sperimentale che un'eccitazione a frequenza $\nu$ possa avvenire solo scambiando con l'esterno energia in multipli di $h\nu$.

La legittimità con cui può essere adottata una descrizione corpuscolare non è però garantita a priori: il cristallo, come qualunque altro stato della materia, presenta solo delle «modalità d'interazione» con l'esterno, oggettive in quanto osservabili e riproducibili. Riteniamo che tali «modalità d'interazione» siano descrivibili in termini di eccitazioni elementari.

Se, con il comportamento di tali eccitazioni elementari, riusciremo a descrivere fenomeni effettivamente osservabili, allora (e solo allora) sarà legittimo attribuire una realtà oggettiva al «fonone», o a oggetti dall'esistenza spesso più effimera quali sono le tante «particelle» che si producono negli acceleratori.

È comprensibile come tale modo di procedere possa indurre dubbi e diffidenze nei «non addetti ai lavori». Ma ciò fa parte di quel processo in cui, all'ampliarsi del campo dell'esperienza, l'idea che abbiamo del mondo, e il significato stesso delle parole che usiamo per descriverlo, mutano e, talora, si evolvono.


## PROBLEMI DEL TERZO MONDO

LE SCIENZE *edizione italiana di* SCIENTIFIC AMERICAN *ha dedicato all'argomento diversi articoli:*

**LE POPOLAZIONI DEI PAESI IN VIA DI SVILUPPO**
di P. Demeny (n. 79)
*Queste popolazioni, che costituiscono quasi i tre quarti della specie umana, continueranno a crescere rapidamente nei prossimi decenni. È necessario accelerare il processo di autoregolazione.*

**TECNOLOGIA E PAESI IN VIA DI SVILUPPO**
di G. Myrdal (n. 79)
*L'introduzione delle tecnologie moderne nei paesi in via di sviluppo richiede notevoli adattamenti per la diversa situazione socioeconomica e culturale e per la diversa disponibilità dei fattori di produzione.*

**IL POTENZIAMENTO DELL'AGRICOLTURA NEI PAESI IN VIA DI SVILUPPO**
di W. D. Hopper (n. 104)
*I paesi poveri potranno raggiungere l'autosufficienza alimentare se la loro agricoltura verrà modernizzata e la loro economia ristrutturata con il trasferimento di risorse tecniche e finanziarie dalle nazioni ricche.*

**IL PIANO DI VALORIZZAZIONE DELLA VALLE DEL GIUBA**
di P. G. Cannata (n. 110)
*Questo piano, frutto di una collaborazione internazionale, è un chiaro esempio di come le nuove strategie di sviluppo debbano tenere conto delle reali esigenze del paese in cui sono attuate.*

**LA LOTTA CONTRO L'EROSIONE DEL TERRENO IN NIGERIA**
di P. G. Cannata e D. Fanciullacci (n. 155)
*Nella regione dell'ex Biafra, dove l'ambiente tropicale è altamente degradato e sovrappopolato, è in fase di realizzazione un progetto che utilizza contemporaneamente strumenti tecnici e di recupero sociale.*

# La memoria negli uccelli che fanno incetta di cibo

*Gli uccelli che nascondono semi per poi recuperarli sembrano avere un'eccellente memoria spaziale: la nocciolaia, per esempio, ricorda anche per diversi mesi dove ha immagazzinato migliaia di provviste*

di Sara J. Shettleworth

In Inghilterra, in inverno, un distributore di cibo ripieno di noccioline è tutto quanto serve per poter osservare il comportamento degli uccelli e dal loro comportamento si può dedurre che questi animali hanno una notevole capacità di memoria, addirittura, sembra, di gran lunga superiore a quella di qualsiasi altro animale che sia stato osservato in laboratorio. Il distributore attrae molti uccelli, tra i quali cinciallegre, cinciarelle e cince bigie, tutte specie di piccole dimensioni e molto vivaci. Le cinciallegre e le cinciarelle si affollano attorno al distributore, mangiando quanto più rapidamente possono. Interrompono il loro pasto solo per scacciare i loro concorrenti. Invece, una cincia bigia giunge come un fulmine, carpisce una nocciolina e fugge via, ma torna quasi istantaneamente per prenderne un'altra. Immagazzina poi le noccioline nelle vicinanze, ciascuna in un posto diverso, fino a quando il distributore è vuoto. Allora comincia ad andare alla ricerca del cibo che ha nascosto.

Negli Stati Uniti sudoccidentali un altro uccello, la nocciolaia di Clark, affine alle ghiandaie e ai corvi, mostra un analogo comportamento: alla fine dell'estate raccoglie i pinoli dei pini, se ne riempie a più riprese la tasca sublinguale, una sacca sotto la lingua, e vola via per parecchi chilometri per andarli a seppellire. Spesso i siti dove avviene il seppellimento si trovano su pendii senza vegetazione, rivolti a mezzogiorno, dove più avanti nel corso dell'anno la neve non raggiungerà mai un notevole spessore. Una nocciolaia può seppellire fino a 33 000 pinoli, formando provviste di quattro-cinque semi ciascuna. Nel corso dell'inverno ritorna e incomincia a scavare le sue migliaia di nascondigli.

Come fanno questi uccelli a trovare le loro provviste? Le cince o le nocciolaie ricordano davvero dove hanno immagazzinato ogni nocciolina, ogni provvista di pinoli? Fino a non molto tempo fa, i ricercatori che osservavano il comportamento in natura degli uccelli che fanno incetta di cibo dubitavano che essi potessero basarsi sulla memoria. Tanto per dirne una, l'uccello dovrebbe davvero avere una memoria capace per ricordare i siti dove ha posto centinaia o migliaia di provviste, raccolte individualmente. La memoria dovrebbe essere, inoltre, duratura. Anche un uccello, come la cincia bigia, che impiega poco tempo ad andare a nascondere il proprio cibo, non lo recupera prima che siano trascorse ore o giorni; un uccello, poi, che impieghi molto tempo a immagazzinare il proprio cibo (per esempio la nocciolaia di Clark), non ritorna al suo bottino per mesi, forse solo al sopraggiungere della primavera, quando deve nutrire con i pinoli la propria prole. Inoltre, in teoria, non è necessaria la memoria per il recupero delle provviste: un uccello, per esempio, potrebbe immagazzinare il cibo di cui ha fatto incetta solo in certi tipi di nascondigli, come quelli sui pendii rivolti verso sud o le cavità nelle cortecce dei tronchi d'albero. Per il recupero, basterebbe che cercasse solo in luoghi di questo genere. Allora potrebbe impiegare la memoria solo per l'area in cui ha immagazzinato il proprio cibo, più che per i singoli nascondigli; all'interno dell'area, poi, potrebbe andare alla ricerca per tentativi ed errori o - come si può immaginare - sollecitato da stimoli quali l'odore.

Gli psicologi che studiano l'apprendimento e la memoria negli animali si stanno, tuttavia, sempre più convincendo che certe specie mostrano specializzazioni adattative, che le rendono particolarmente capaci di apprendere e ricordare cose che rivestono per esse una particolare importanza. Tra gli esempi ben noti, si ricorda la capacità di molti uccelli di imparare i canti della propria specie, la capacità di ratti e altri animali di ricordare localizzazioni spaziali, la capacità delle api di ricordare la posizione dei fiori e, infine, la capacità di molti animali di imparare a evitare gli alimenti nocivi. Se poi, davvero, gli uccelli che immagazzinano il cibo sono in grado di ricordare un gran numero di nascondigli per lunghi periodi di tempo, tale loro memoria potrebbe essere un altro esempio di specializzazione adattativa, che renderebbe gli uccelli capaci di recuperare le loro provviste con notevole efficienza. Un uccello che potesse ricordare dove ha nascosto il cibo compirebbe un minor numero di errori per recuperarlo e spenderebbe meno tempo ed energia degli uccelli che, invece, compiono le loro ricerche in modo del tutto casuale. Studi recenti hanno fornito risultati indicativi, secondo cui almeno alcune specie che immagazzinano il cibo ricordano i siti dove hanno nascosto il loro bottino.

Negli ultimissimi anni, John R. Krebs e collaboratori dell'Università di Oxford hanno accumulato alcune tra le prove più particolareggiate della capacità degli uccelli di ricordare i siti esatti in cui immagazzinano il loro cibo. Questi studiosi hanno incominciato con il cercare prove che confermassero che le cince bigie nel loro ambiente naturale recuperano il cibo nascosto. A questo scopo, alcuni esemplari furono addestrati a dirigersi verso distributori di semi di girasole, posti nei territori di questi uccelli nello Wytham Wood, vicino a Oxford. Richard J. Cowie, Krebs e David F. Sherry rivestirono quindi tali semi con una sostanza radioattiva innocua. (Gli uccelli rimuovono il rivestimento esterno prima di mangiare il seme vero e proprio.) La radioattività permise ai tre ricercatori, attrezzati con un contatore a scintillazione portatile, di trovare i semi che gli uccelli avevano nascosto e quindi di controllarli ogni certo numero di ore per vedere quando scomparivano.

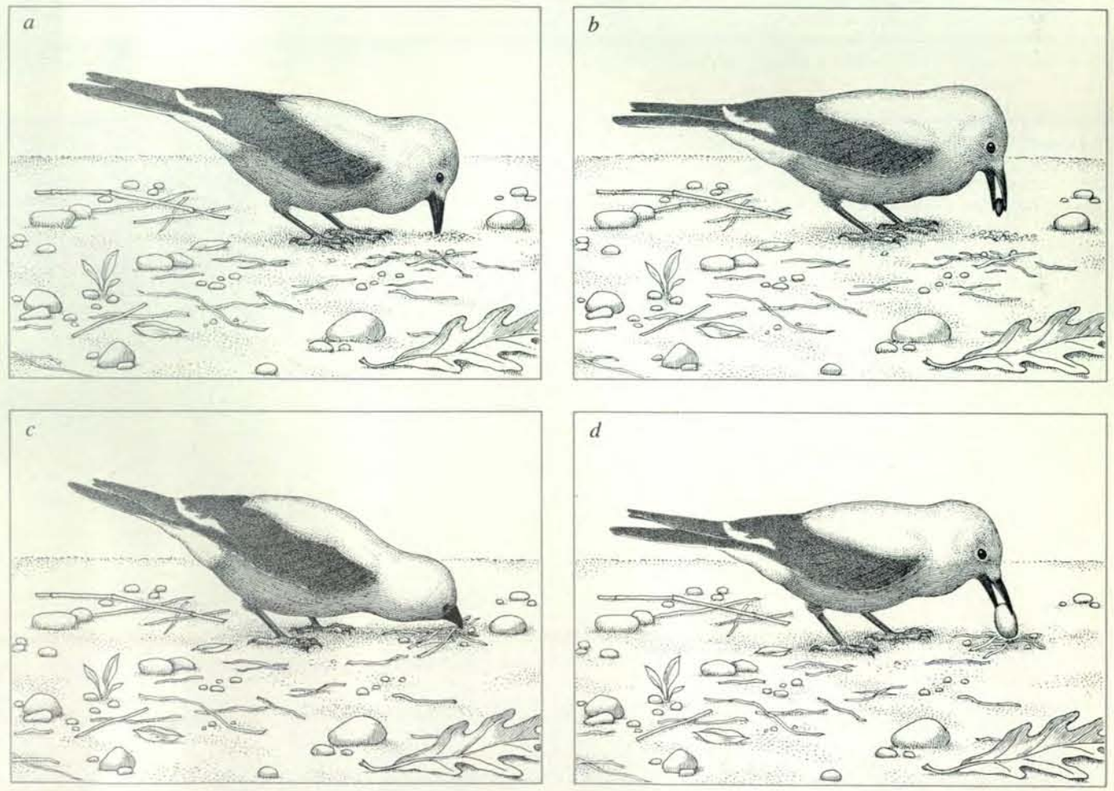


Naturalmente i semi potrebbero essere mangiati da altri uccelli o da roditori come i topi o i toporagni. Cowie, Krebs e Sherry pensarono allora che, se le cince bigie utilizzavano la memoria per recuperare le loro provviste, i semi sarebbero scomparsi più rapidamente dai siti dove esse li avevano immagazzinati che da altri «falsi siti», predisposti dai ricercatori. E ciò, difatti, avvenne. I semi che le cince avevano messo da parte furono consumati più in fretta dei semi che i ricercatori avevano posto in siti analoghi, ciascuno a un metro di distanza da una provvista fatta dagli uccelli. Nello Wytham Wood le provviste naturali, in inverno, venivano consumate in un giorno o due. L'ultima delle false provviste fu esaurita poco dopo, il che suggerisce che la predazione è, comunque, un fatto comune.

Una tattica difensiva che le cince bigie a quanto pare adottano contro gli animali che predano le loro scorte consiste nel disperdere i semi invece che seppellirli a mucchi. In media, la distanza tra i vari nascondigli che le cince bigie avevano allestito nello Wytham Wood era di diversi metri. Sherry, Mark Avery e Allen Stevens prepararono false provviste di semi di girasole e trovarono che quei semi duravano tanto più a lungo quanto più essi erano distanziati, fino a una distanza di circa sette metri. Le scorte più distanti ancora non erano invece più sicure di quelle distanti sette metri. Almeno nello Wytham Wood le cince bigie davano l'impressione di aver raggiunto, per le scorte, la distanza ottimale.

Dopo aver ricavato delle prove che le cince bigie in natura recuperano le loro scorte, Krebs e collaboratori cominciarono a compiere osservazioni in laboratorio. Anche in cattività le cince bigie, dotate di una ciotola con dei semi e di adatti siti di immagazzinamento, formano le loro provviste rapidamente. I tre ricercatori (Krebs, Sherry e Cowie) misero a loro disposizione cassette contenenti muschio, delle dimensioni di un metro di lato e dove si potessero nascondere i semi. Fu chiaro così che le cince si ricordavano benissimo dove avevano messo le provviste, anche dopo ventiquattr'ore. In questi esperimenti, però, il comportamento delle cince veniva registrato solo come visite ad aree relativamente ampie della superficie della cassetta. Inoltre, nel tentativo di vedere se quegli uccelli si servivano davvero solo della memoria nel corso degli esperimenti, i semi nascosti venivano rimossi dalla cassetta prima che essi vi ritornassero. Così, era difficile dire con quale precisione gli uccelli riuscivano a localizzare i siti dove avevano posto le loro provviste; era parimenti difficile identificare gli errori che essi potevano compiere.

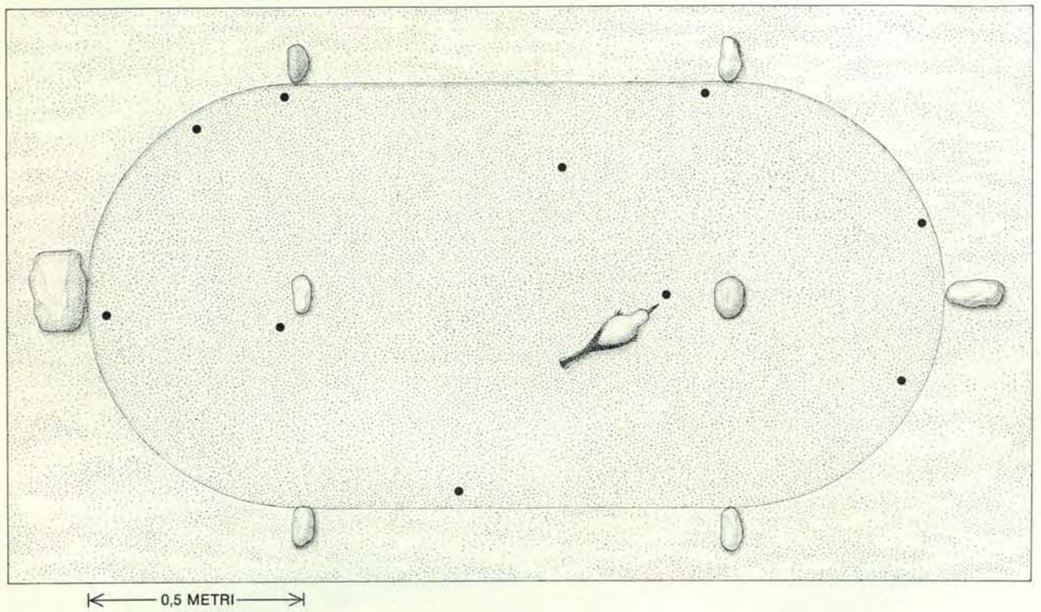
Quando una borsa di studio Guggenheim mi permise di trascorrere un anno sabbatico a Oxford, il mio interesse per le specializzazioni adattative dell'apprendimento e della memoria mi spinse verso lo studio delle cince bigie. Krebs e io progettammo alcuni esperimenti per analizzare in maggior dettaglio l'evidente memoria di questi uccelli ed eravamo particolarmente interessati a scoprire se il cibo immagazzinato più di recente era anche quello che essi recuperavano per primo. Questo «effetto del recente» compare spesso nei test sulla memoria e

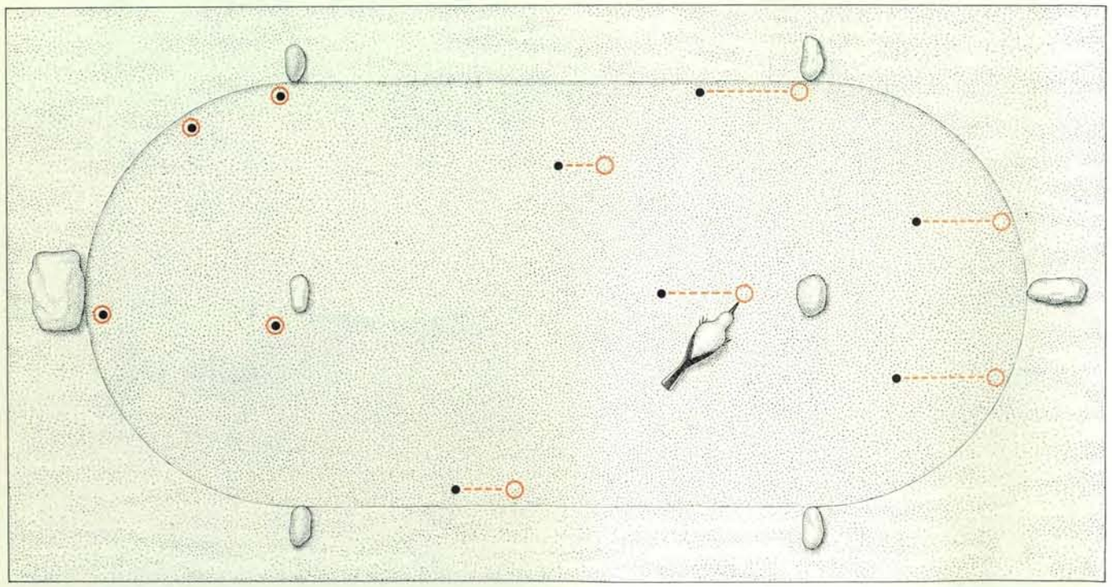



**L'immagazzinamento di cibo da parte di una nocciolaia di Clark è illustrato in una sequenza di disegni basati sulle osservazioni compiute da Stephen B. Vander Wall della Utah State University. La nocciolaia conficca il becco nel suolo per smuoverlo *(a)*, quindi vi inserisce un seme di conifera *(b)* e vi raccoglie sopra terriccio ed erba *(c)*, che trattiene con un sasso *(d)*. Questo serve solo a mimetizzare il sito fino a quando il tempo meteorologico cela ogni traccia del nascondiglio. L'intero procedimento richiede dai 10 ai 20 secondi. In natura, la nocciolaia riesce ad accumulare fino a 33 000 pinoli in migliaia di nascondigli, che ne contengono da quattro a cinque ciascuno e coprono una vasta area. Essa fa incetta da settembre a novembre e si basa sulla memoria per andare a ritrovare le scorte durante il resto dell'anno.**

vi sono validi motivi per prevederlo anche nel caso di incetta di cibo: se le provviste sono destinate a finire in bocca ai saprofagi a una velocità costante, è chiaro che, in un qualsiasi momento, le provviste più recenti sono anche quelle alle quali gli uccelli che le hanno raccolte hanno più facile accesso. Cercarle per prime costituisce, dunque, per questi animali la miglior scelta.

In un vasto locale, Krebs e io ponemmo pezzi di rami d'albero, in ognuno dei quali avevamo praticato fori delle dimensioni giuste per inserirvi semi di canapa. Scavammo in tutto un centinaio di fori e ciascuno venne ricoperto da un lembo di tela, che l'uccello doveva sollevare per poter deporre il seme o prelevarne uno.



**La prova che le nocciolaie di Clark fossero dotate di memoria è emersa da esperimenti realizzati da Vander Wall. Qui, una nocciolaia viene lasciata libera di nascondere dei semi *(pallini neri)* in un'area ovale sul fondo sabbioso di una voliera; a ogni estremità dell'ovale sono state poste quattro pietre *(disegno in alto)*. L'ovale è stato allungato di 20 centimetri e le pietre a destra sono state spostate di conseguenza *(disegno in basso)*. Due giorni dopo l'uccello ha potuto cercare le proprie scorte: a sinistra *(cerchietti in colore)* ha scavato con successo; a destra, ha compiuto degli errori di circa 20 centimetri. È evidente che si basava sul ricordo della posizione delle pietre vicino alle quali aveva nascosto i semi. Gli errori compiuti al centro dell'ovale fanno pensare, invece, che stimasse la posizione di alcune provviste riferendosi ad ambedue le serie di pietre.**

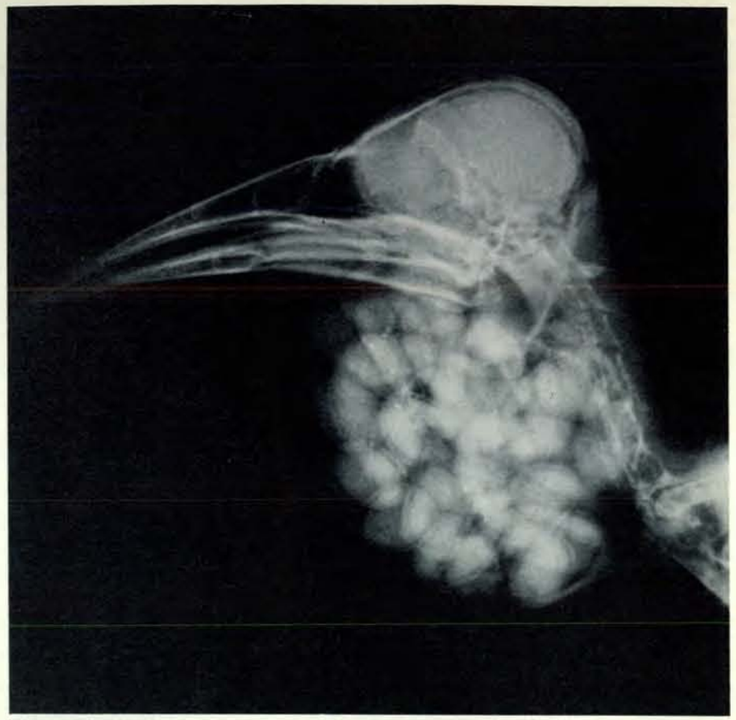


Quel locale fu sede di tutta una serie di esperimenti. Nel nostro primo esperimento, per ogni prova abbiamo lasciato che una cincia bigia immagazzinasse dodici semi che si era procurata da una ciotola posta sul pavimento al centro del locale. Ogni foro poteva contenere solo un seme, quindi i vari semi vennero posti in altrettanti fori diversi. Una volta nascosti i semi, tenemmo l'uccello fuori dalla stanza per circa due ore e mezzo. Levammo la ciotola con i semi e quindi facemmo entrare di nuovo l'animale che subito andò alla ricerca dei semi che aveva nascosto. Se avesse dovuto ricercare casualmente tra cento fori, avrebbe dovuto esaminare una media di circa otto fori per trovare un seme. In realtà l'efficienza delle cince si rivelò molto più elevata. In media, ogni soggetto faceva due errori per seme. All'inizio di un test di recupero, un uccello trovava direttamente tre o quattro semi senza mai cercare in fori vuoti.

Poteva darsi che, per trovare i semi, invece di utilizzare la memoria, le cince si servissero dell'olfatto o di qualche altro mezzo al quale non avevamo pensato. L'olfatto sembrava improbabile; tale senso è generalmente poco sviluppato negli uccelli. Ancora una volta, per saggiare queste possibilità, lasciammo che una cincia immagazzinasse dodici semi e quindi glieli spostammo in altri fori (fori che essa aveva utilizzato in precedenti prove). In queste circostanze, i soggetti reintrodotti nel locale fecero circa sei errori per seme e, alla fine, guardando in più fori del solito, riuscirono a recuperare i due terzi dei semi.

Questi risultati indicano che le cince bigie non scoprivano i semi nascosti con l'olfatto o mediante qualche mezzo analogo. D'altra parte, la loro prestazione nel primo esperimento non significa necessariamente che si ricordino dove hanno immagazzinato esattamente un seme su cento fori. In una certa misura, ogni uccello del laboratorio sviluppò una preferenza per fori nei quali depositava semi con particolare frequenza; con ogni probabilità, quando cercava cibo, ispezionava per primi i fori preferiti. L'operazione di recupero dei semi da questi poteva essere dovuta all'abitudine dell'uccello - e a nulla di più - di visitare quei fori, sia per depositarvi semi, sia per prelevarli. In verità, questa tendenza probabilmente spiega perché, quando spostammo i semi nascosti, la prestazione delle cince fu migliore di una prestazione puramente dovuta al caso.

Un più attento esame dei nostri dati mostrò che una cincia non andava semplicemente agli stessi fori a ogni test. Anche i fori che preferiva di più venivano ispezionati con maggiore probabilità, durante il recupero dei semi, se la cincia vi aveva effettivamente posto un seme. Il problema era, ciononostante, di fondamentale importanza: nei convenzionali test di laboratorio sulla memoria, lo sperimentatore fornisce all'animale l'informazione che deve cercare di ricordare. Nelle nostre prove, invece, furono gli uccelli stessi a fornire l'informazione quando immagazzinarono i semi nei siti prescelti. Qualunque tendenza che potessero avere a guardare dentro gli stessi fori, in visite successive al locale dove si svolgevano le prove, li avrebbe indotti a comportarsi come se la loro memoria nei riguardi delle scorte immagazzinate fosse buona.

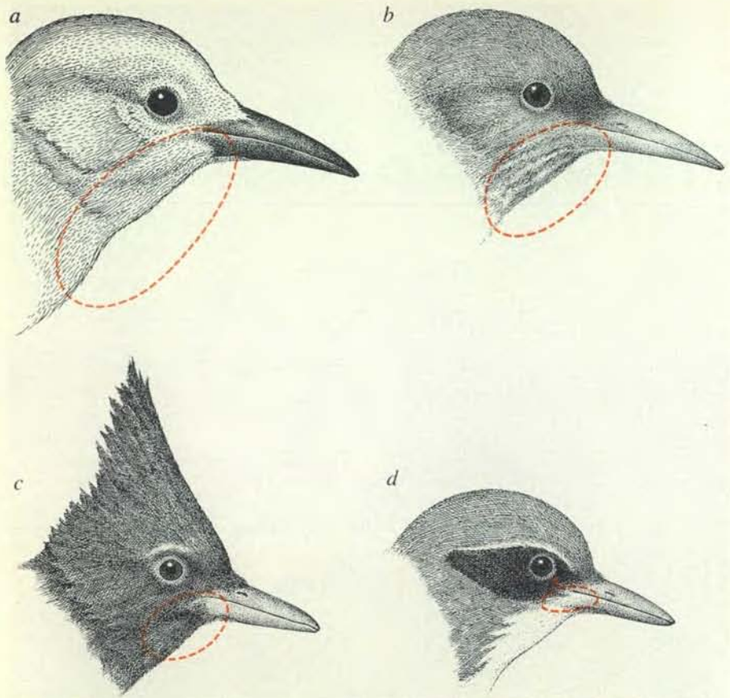


**La tasca sublinguale della nocciolaia di Clark è una sacca situata sotto alla lingua nella quale questo uccello trasporta i semi. Si tratta di una specializzazione anatomica che rende la nocciolaia perfettamente idonea a trasportare e a immagazzinare semi. Quando Vander Wall ha scattato questa fotografia ai raggi X, la tasca conteneva 38 semi con un peso netto di 30,6 grammi.**

Pertanto, il nostro esperimento successivo fu programmato in modo da far lavorare la memoria degli uccelli contro la loro tendenza a preferire certi siti di immagazzinamento. Di nuovo a ogni uccello vennero fatti immagazzinare dei semi, poi lo si fece uscire dal locale, per riammettervelo circa due ore più tardi. Nella seconda sessione di prove, invece di far cercare semi all'animale, lo lasciammo accumulare altri semi. Perché lo facesse, lasciammo soltanto una ciotola piena di semi nella camera. Secondo noi, se le cince ricordavano quali fori contenevano già semi, li avrebbero evitati nell'immagazzinare il secondo lotto. O, all'inverso, se ispezionavano gli stessi fori ogni volta che visitavano il locale, molti fori che ispezionavano durante la seconda visita sarebbero già stati occupati.

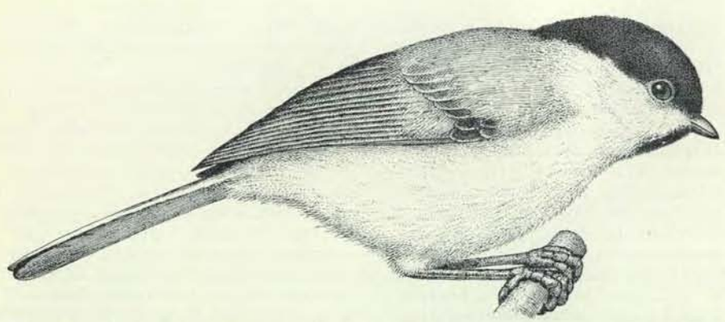
I risultati furono chiari. Quando immagazzinavano un secondo lotto di semi, gli uccelli non ispezionavano quasi mai fori che già ne contenevano uno. Quando però avevano fame e non c'erano più semi da immagazzinare, mostravano la loro abituale efficienza nel recuperare i semi dai fori. Fu interessante notare che un uccello che deve immagazzinare un secondo lotto di semi spesso sceglie fori che si trovano vicino a quelli dei quali si era servito per immagazzinare il primo lotto. Sembra, dunque, che si ricordi ed eviti singoli fori e non zone della stanza o rami particolari.

Dopo aver immagazzinato due lotti di semi, lasciammo che l'uccello li recuperasse tutti. In questo modo riuscimmo a saggiare l'effetto del recente di cui abbiamo parlato. Se le cince lo possedessero, dovrebbero recuperare i semi del secondo lotto prima o più spesso di quelli del primo. I risultati che abbiamo ottenuto suggeriscono che un effetto del genere esiste, ma non è molto spiccato. Se però la memoria dei siti di immagazzinamento si attenua lentamente, due episodi di incetta del cibo dovrebbero essere distanziati da più di due ore perché l'effetto del recente possa essere chiaro.




**La specializzazione anatomica di quattro uccelli che vivono negli Stati Uniti sudoccidentali corrisponde al grado variabile di dipendenza dalle scorte di cibo. La nocciolaia di Clark *(a)* ha il maggior grado di dipendenza ed è la più specializzata: oltre a una tasca sublinguale possiede un lungo becco affilato con cui apre le pigne per ricavarne i pinoli. Inoltre, è un robusto volatore che può compiere lunghi tragitti con un carico di semi. La ghiandaia dal capo azzurro *(b)* e la cianocitta di Steller *(c)* hanno anatomia e comportamento intermedi. Trasportano semi nel loro esofago estensibile e non in una tasca sublinguale specializzata. L'afelocoma azzurra *(d)* è la meno dipendente dalle scorte di semi e possiede l'anatomia meno specializzata. Può portare i semi solo in bocca e nel becco. La capacità di trasportare semi viene indicata da una linea tratteggiata in colore.**

Un altro risultato è emerso dagli esperimenti. È ovvio che un buon ricercatore di cibo immagazzinato raramente dovrebbe ispezionare siti vuoti. Inoltre, dovrebbe essere in grado di ricordare dove è già stato per recuperare cibo, in modo da evitare di ritornare sui suoi passi. Ratti di laboratorio, messi di fronte al compito di raccogliere cibo da un certo numero di siti posti in un ampio labirinto, dimostrarono di avere quest'ultima capacità. Così anche le cince bigie. Negli esperimenti che Krebs e io effettuammo, i fori che un uccello ispezionava nell'operazione di recupero dei semi erano ispezionati una seconda volta molto meno di quanto si poteva prevedere se l'uccello non avesse ricordato ed evitato tali siti. Negli esperimenti svolti da Sherry, gli uccelli tendevano, infatti, a non ispezionare aree di muschio dove già avevano cercato semi. Pertanto, una cincia bigia in cerca di cibo immagazzinato si serve di due tipi di informazioni: ricorda dove si trovano le provviste e al tempo stesso memorizza i siti che ha già ispezionato.



**La cincia bigia vive in Gran Bretagna e in altri paesi europei. Immagazzina i semi se sono disponibili in quantità maggiori di quelle che le servono al momento come alimento. Il suo comportamento tipico la porta a ritornare sulle scorte che ha nascosto entro poche ore o al massimo pochi giorni.**

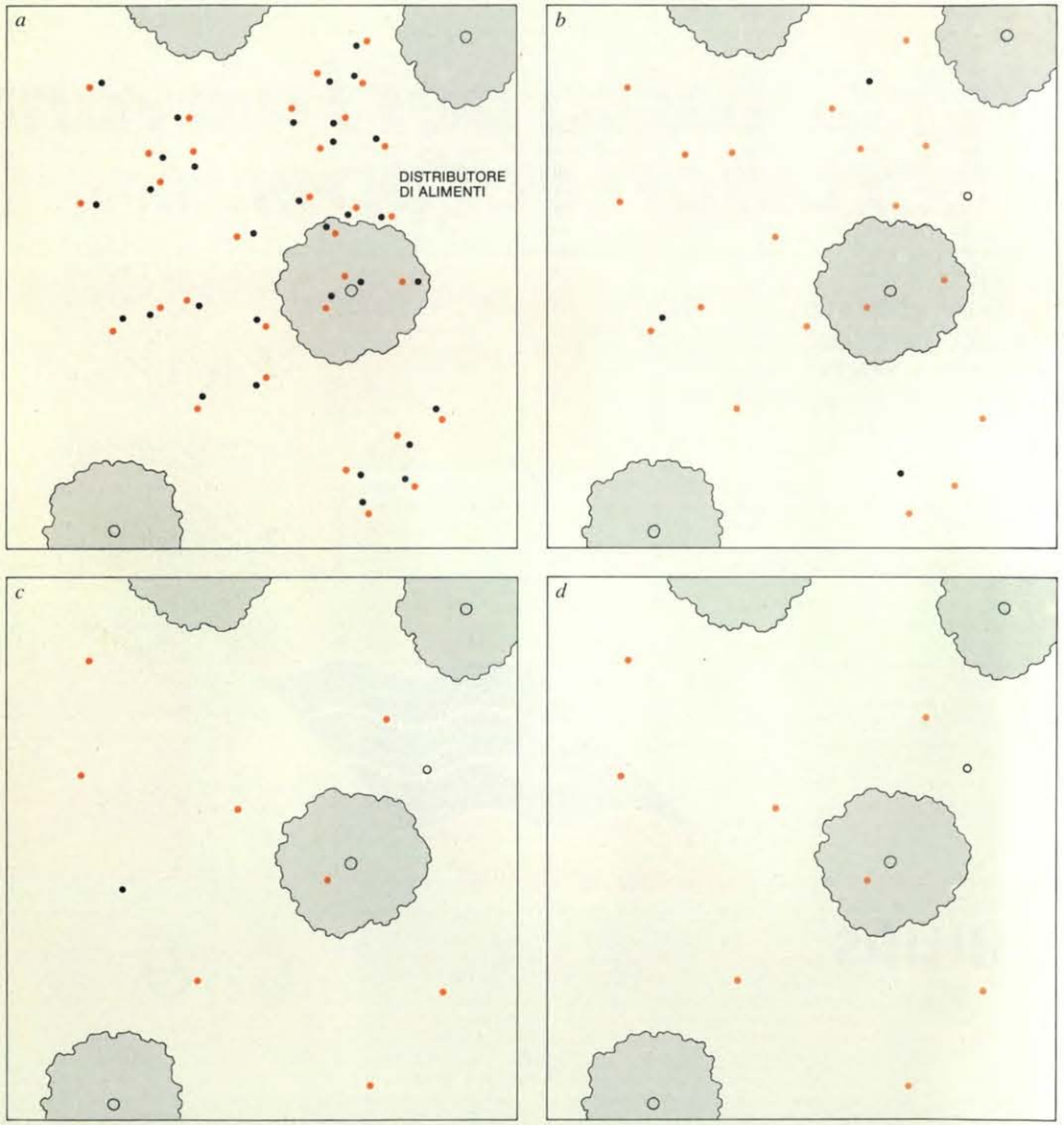
Se l'informazione sui siti di immagazzinamento viene archiviata nel cervello della cincia, la memoria di quest'ultima dovrebbe essere soggetta a una peculiarità del sistema nervoso che sembra essere presente in molti uccelli. In numerosi vertebrati (tra cui l'uomo) le informazioni relative alla metà sinistra del campo visivo provengono da ambedue gli occhi, ma sono dirette alla metà destra del cervello e viceversa. La situazione è leggermente diversa in uccelli come la cincia bigia, in cui gli occhi sono situati ai lati del capo e ogni occhio ispeziona un campo visivo distinto. La ricostruzione delle vie percorse dall'informazione nel cervello di questo uccello e i risultati di esperimenti realizzati in prevalenza con piccioni fanno pensare che in questi animali l'informazione che raggiunge il cervello attraverso un occhio venga immagazzinata in primo luogo nella metà del cervello che si trova sul lato opposto del capo. In breve, le cince bigie hanno uno scarso trasferimento interoculare per molti tipi di informazione o non ne hanno affatto. Sherry, in collaborazione con Krebs e Cowie, si servì di questo fatto per interpretare un interessante comportamento che le cince bigie mostrano quando immagazzinano un seme. Dopo aver dato un colpetto al seme o averlo fatto penetrare nel foro, l'uccello immediatamente piega la testa all'indietro prima da un lato e quindi dall'altro, come se stesse guardando il sito di immagazzinamento o i punti di riferimento vicini, prima con un occhio e poi con l'altro.

La cincia compie questi movimenti della testa per poter archiviare l'informazione visiva sui siti di immagazzinamento in ambedue le metà del cervello? Per rispondere a questo interrogativo Sherry fece immagazzinare alle cince semi di girasole, dovunque volevano in una cassetta piena di muschio e facendo in modo che un loro occhio fosse coperto da una mascherina traslucida. Se in un secondo tempo, gli stessi soggetti, sempre con l'occhio coperto, dovevano ricercare i semi, li ritrovavano normalmente. Se, invece, la mascherina veniva spostata sull'altro occhio, in modo che le ricerche dovevano essere effettuate con l'occhio che non aveva visto i siti di immagazzinamento, la cincia si comportava come se non si ricordasse dove erano i semi. Se nelle cince non si verifica affatto trasferimento interoculare di informazione, o esso è molto scarso, il loro comporta-

mento è un'ulteriore prova che esse, per recuperare le loro provviste, fanno affidamento sulla memoria.

Che dire di altri uccelli che immagazzinano il cibo? Negli ultimissimi anni, diversi ricercatori hanno studiato uccelli come la nocciolaia di Clark, che costituisce, per l'inverno, migliaia di riserve. Anche per questa specie, sia in natura sia in laboratorio sono state cercate prove della presenza di una memoria. Naturalmente uno studio sul campo per cercare di scoprire, per esempio, se una nocciolaia di Clark recupera o no le proprie provviste, è molto difficile da compiere a causa del numero elevato di nascondigli e del lungo intervallo di tempo che intercorre tra l'immagazzinamento e il recupero. Ben pochi sarebbero disposti ad aspettare parecchi mesi vicino al nascondiglio di una nocciolaia, sperando di riuscire a sapere prima o poi se verrà di nuovo aperto dallo stesso uccello che vi aveva posto le sue riserve.

Diana Tomback dell'Università del Colorado a Denver ha sostenuto, però, che si potrebbero imparare molte cose dal modo in cui le nocciolaie di Clark ricercano le loro provviste, e ha sfruttato




**La scomparsa delle scorte di cibo dallo Wytham Wood vicino a Oxford è stata la prima conferma del fatto che le cince bigie utilizzano la memoria per recuperare il cibo nascosto. R. J. Cowie, J. R. Krebs e D. F. Sherry dell'Università di Oxford hanno rivestito 20 semi di girasole con una sostanza radioattiva, quindi con un contatore a scintillazione portatile hanno trovato dove una cincia bigia li aveva nascosti *(a, pallini neri)*. A un metro di distanza da ogni nascondiglio i ricercatori ne fissarono uno falso *(pallini in colore)*. Dieci ore dopo *(b)*, 19 falsi nascondigli erano rimasti indisturbati, mentre dei nascondigli dell'uccello solo tre non vennero toccati. Vennero compiuti ulteriori controlli dopo 19 *(c)* e 30 ore *(d)*.**



il fatto che questi uccelli lasciano una documentazione delle loro ricerche sotto forma di segni dei loro becchi sulla neve o sulla terra che scavano. Si possono, inoltre, identificare i successi di queste ricerche nella presenza di rivestimenti di pinoli vicino alle buche dalle quali questi semi sono stati prelevati. Se le nocciolaie ricercassero i loro bottini a casaccio, i nascondigli esplorati con successo e i gruppi di quelli esplorati a vuoto dovrebbero essere più o meno uniformemente distribuiti nelle aree dove esse hanno svolto le loro ricerche. Al contrario, Diana Tomback ha trovato che quelli esplorati a vuoto erano addensati attorno agli altri. Un simile quadro suggeriva che le nocciolaie non dovevano aver cercato i loro bottini per tentativi ed errori. Inoltre, all'inizio della primavera (cioè prima che i roditori avessero scoperto molte delle provviste), circa i due terzi dei nascondigli erano stati esplorati con successo, una quantità di gran lunga superiore a quella che ci si aspetterebbe nel caso di una ricerca condotta a casaccio.

Le osservazioni di Diana Tomback non dimostrano in modo inequivocabile che le nocciolaie fossero in grado di ricordare dove si trovavano le loro provviste. Per esempio, potevano averle scoperte con l'olfatto (anche se, di nuovo, questa possibilità è parsa improbabile). O, in alternativa, potevano averle cercate principalmente nelle località del tipo di quelle dove con tutta probabilità sarebbe stato facile trovarle. Recenti studi di laboratorio, effettuati da Stephen B. Vander Wall presso la Utah State University sembrano escludere queste possibilità. Vander Wall ha studiato quattro nocciolaie di Clark in cattività in una grande uccelliera all'aperto. Due di questi esemplari subito nascosero i semi nel terreno sabbioso del recinto; gli altri due no. Malgrado questo, anche gli uccelli che non avevano accumulato riserve scavarono nella sabbia con il becco e mangiarono i semi sepolti che erano riusciti a scoprire. Vander Wall fece in modo che i due uccelli che costituivano scorte potessero seppellire i loro semi nel recinto; quindi lasciò liberi tutti e quattro gli esemplari di cercare i semi nascosti.

I risultati fornirono una prova sensazionale dell'esistenza di una memoria in questi animali. Quando le nocciolaie che avevano fatto incetta di semi scavarono nella sabbia per trovarli, il 70 per cento dei loro tentativi andò a buon frutto; quando, invece, si misero all'opera gli uccelli che non avevano allestito scorte solo il 10 per cento dei loro tentativi ebbe successo. Senza dubbio, un successo del 10 per cento è sostanzialmente superiore a quello che ci si potrebbe aspettare da un uccello che compie i suoi tentativi completamente a caso. D'altra parte, le due nocciolaie che nascondevano i semi preferivano farlo vicino a oggetti sporgenti, tronchi oppure pietre poste nella voliera; tutti e quattro gli uccelli il più delle volte cercavano proprio attorno a questi oggetti. Una simile tendenza spiega il successo ottenuto nella ricerca anche dagli uccelli che non avevano fatto provvista di semi.

Forse la presenza di oggetti sporgenti vicini serve da stimolo per la memoria quando gli uccelli tentano di trovare le loro scorte. Per comprovare questa ipotesi, Vander Wall coprì con un foglio di plastica il fondo della voliera, lasciando scoperta solo un'area ovale in cui la plastica era stata tagliata. A ogni estremità di quest'area pose quattro oggetti, di grandi dimensioni, come quattro pezzi di roccia. Le nocciolaie furono lasciate libere di allestire le loro scorte dove volevano nell'ovale. A questo punto, una delle estremità dell'ovale venne estesa di 20 centimetri e anche i quattro oggetti ivi localizzati furono spostati in direzione dell'allungamento per cui vennero anch'essi a trovarsi a 20 centimetri dalla posizione originale.

Le nocciolaie erano ancora in grado di trovare le loro provviste in corrispondenza dell'estremità modificata così come trovavano quelle all'altra estremità non modificata, oppure i loro tentativi andavano a vuoto per 20 centimetri? Si verificò proprio questo secondo caso. L'entità del successo in corrispondenza dell'estremità non modificata dell'ovale fu elevata e si abbassò, invece, verso l'estremità modificata. La maggior parte dei sondaggi in corrispondenza dell'estremità modificata avvenne entro pochi centimetri dal punto ritenuto «giusto» rispetto alla posizione alterata del grosso oggetto più vicino. Gli uccelli compivano errori anche vicino al centro dell'ovale, ma erano errori meno marcati. Questo risultato fa pensare che i siti per le provviste erano stati prescelti con riferimento ad ambedue le serie di oggetti.

Per qualunque specie che disperda le proprie provviste e si basi sulla memoria per il loro recupero, ci si può chiedere se la memoria venga sfruttata anche in altro senso. In altre parole, le specie che immagazzinano il cibo sono in generale avvantaggiate nel ricordare dove sono le cose, rispetto alle altre specie? Il tentativo di rispondere a un interrogativo del genere richiederebbe di sottoporre gli animali a test di memoria spaziale che non comportino l'immagazzinamento di cibo. Test del genere non sono stati ancora effettuati con uccelli che immagazzinano il cibo, ma il solo riflettere sul modo in cui potrebbero essere impostati mette in luce alcuni punti importanti.

Tra le tecniche di laboratorio utilizzate dagli psicologi per studiare la memoria degli animali, l'analogo più vicino all'operazione di far incetta di cibo è un esperimento a risposta ritardata, in cui viene presentato a un animale qualche stimolo visivo, per esempio una luce lampeggiante o una figura geometrica. Quindi viene offerta una scelta tra questo stimolo originario e un altro stimolo e si ricompensa l'animale per aver scelto quello che ha notato per primo. Se passano alcuni secondi o addirittura alcuni minuti tra la prima visione e la seconda, la prestazione di un ratto di laboratorio oppure di un piccione scende tipicamente al livello


## ARTE

LE SCIENZE *edizione italiana di* SCIENTIFIC AMERICAN

*ha dedicato all'argomento diversi articoli:*

**L'olografia nel campo del restauro**
di F. Gori e G. Urbani (n. 74)

**L'origine dell'ambiguità nelle opere di Maurits C. Escher**
di M. L. Teuber (n. 75)

**Pieter Bruegel il Vecchio e la tecnica del Cinquecento**
di H. A. Klein (n. 117)

**La conservazione della pietra**
di K. L. Gauri (n. 120)

**I disegni preistorici tracciati sul terreno in Perù**
di W. H. Isbell (n. 124)

**L'«infrarosso colore» nell'indagine dei dipinti**
di M. Matteini, A. Moles e P. Tiano (n. 142)

**La statica dall'arte alla scienza**
di S. Clara Roero (n. 150)

**Norme architettoniche nella Cina del XII secolo**
di E. Glahn (n. 155)

**L'architettura di Christopher Wren**
di H. Dorn e R. Marck (n. 157)

**L'acustica dei piani armonici di violino**
di C. Maley Hutchins (n. 160)

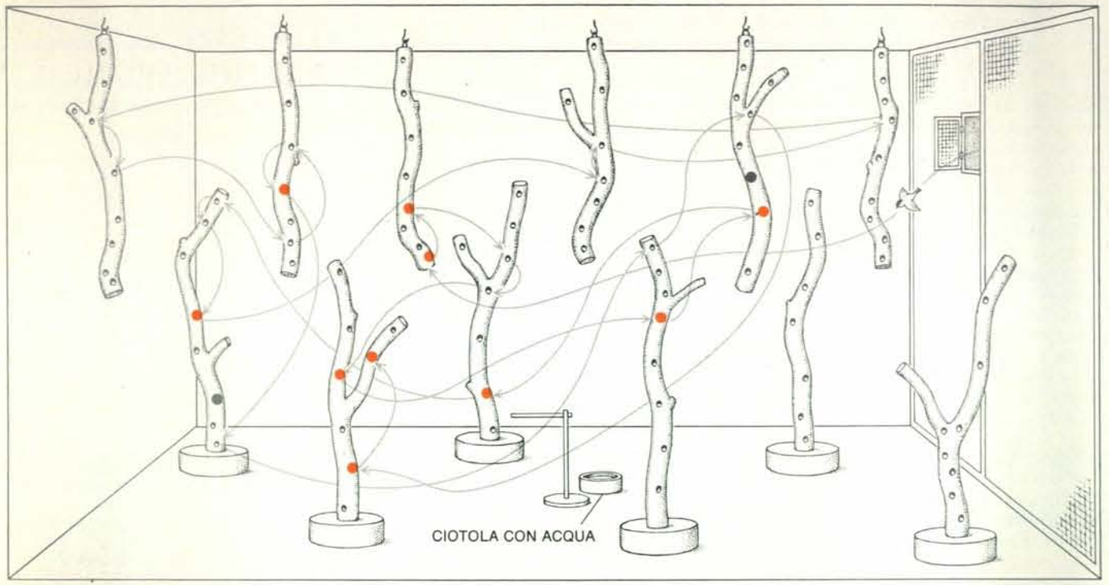
**Conservazione e restauro**
di P. Parrini (n. 161)

**Intarsi rinascimentali: l'arte della geometria**
di A. Tormey e Y. Farr Tormey (n. 169)

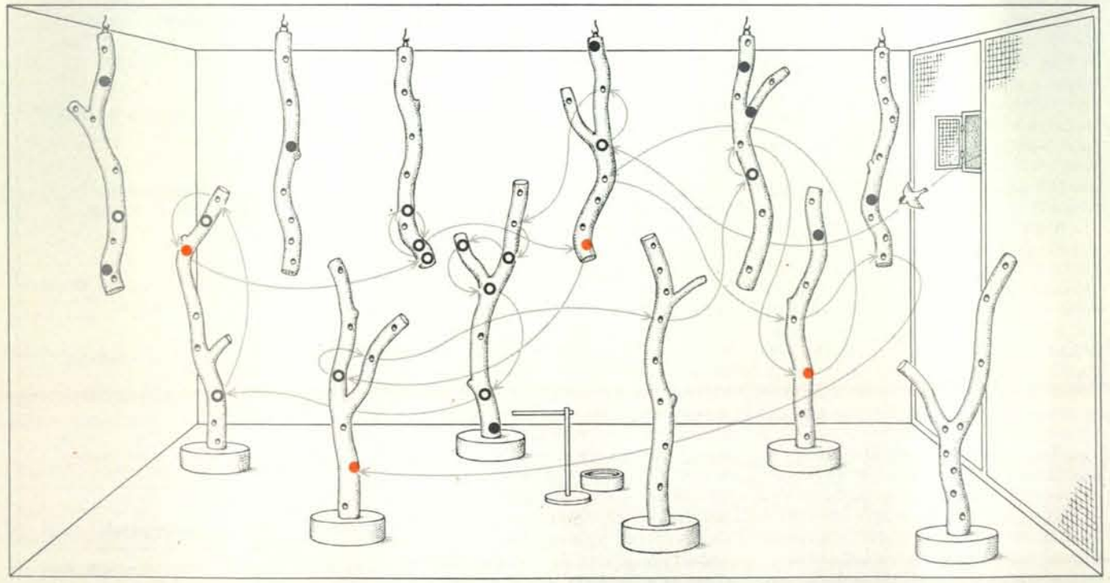
di un comportamento del tutto casuale.

Di recente, Donald M. Wilkie e Russel Summers dell'Università della Columbia britannica hanno addestrato dei piccioni basandosi su un procedimento che è una variante di quello della risposta ritardata e richiede memoria spaziale. Tale lavoro illustra alcune differenze che esistono tra studi di laboratorio convenzionali sulla memoria animale e studi sugli uccelli che fanno incetta di cibo. Nell'esperimento di Wilkie e Summers, un piccione veniva messo di fronte a una serie di nove dischi bianchi disposti a formare una matrice di tre per tre, che potevano anche essere illuminati individualmente da dietro. All'inizio di ogni prova, un disco veniva illuminato per breve tempo. Pochi secondi dopo veniva illuminato di nuovo, insieme con un altro disco o con diversi altri. Il piccione veniva ricompensato (riceveva un poco di cibo) se beccava il disco che era stato illuminato due volte.

Il compito del piccione sembra simile a quello che deve affrontare un uccello che immagazzini cibo. Dopo tutto, quest'ultimo è esposto a stimoli visivi che provengono dal luogo dell'immagazzinamento e, per poter recuperare il cibo, deve selezionare il sito preferendolo ad analoghi siti

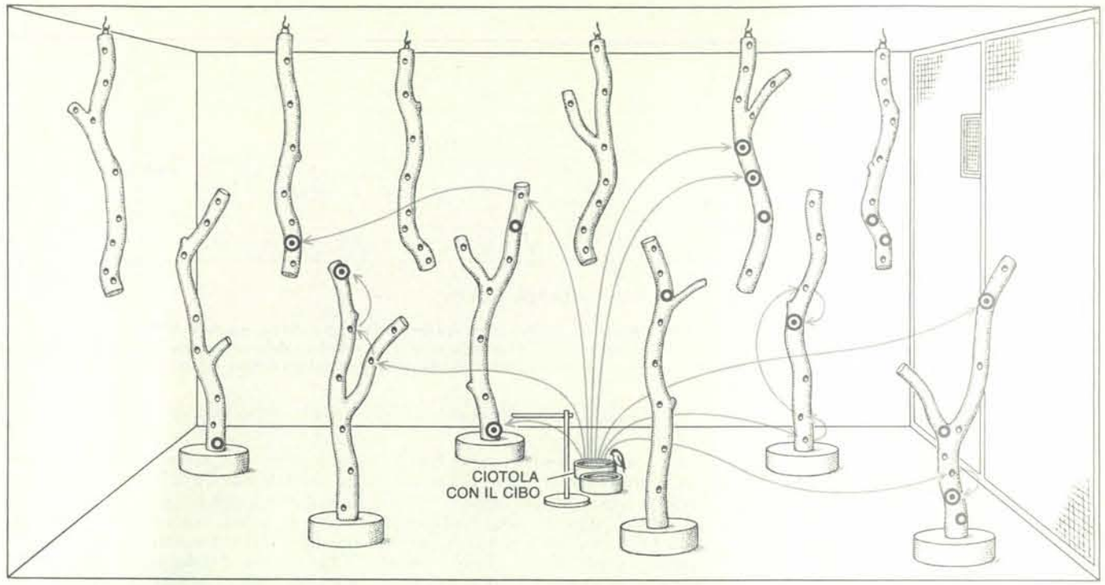



**L'autore e Krebs hanno compiuto 4 esperimenti per provare la memoria delle cince bigie. In ogni caso, uno di questi uccelli ha immagazzinato semi in alcuni dei circa 100 fori che gli sperimentatori avevano praticato nei rami d'albero posti nel locale del laboratorio. Qui, una cincia bigia ha nascosto dodici semi *(pallini neri e in colore)*: due ore e mezzo dopo, ha trovato dieci dei dodici semi *(pallini in colore)* ispezionando solo 24 fori.**

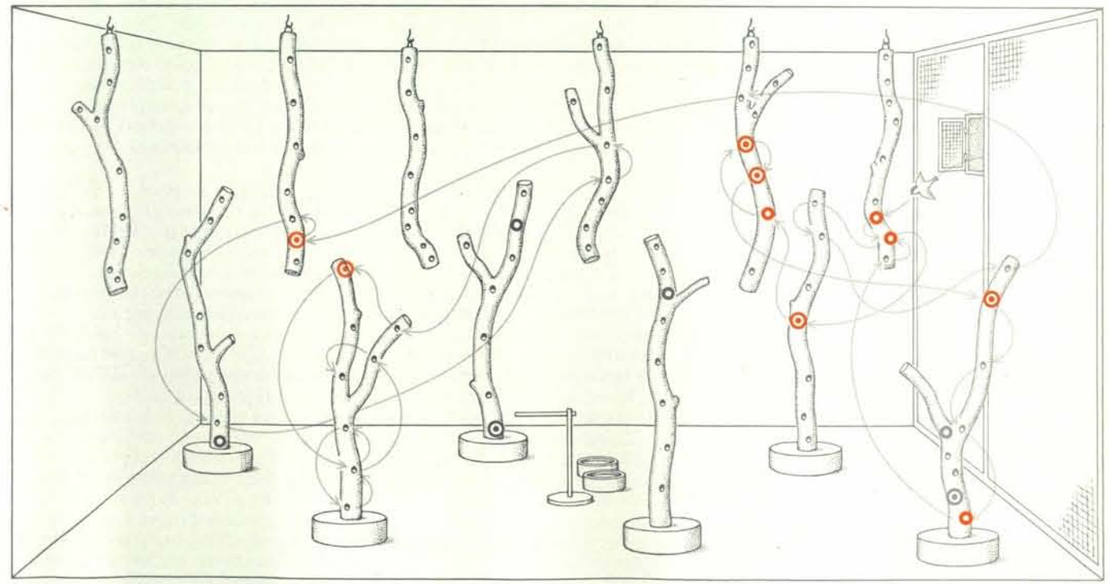


**In un secondo esperimento, una cincia bigia ha immagazzinato 13 semi *(cerchietti neri)*, che dopo sono stati spostati altrove *(pallini neri e in colore)* da Krebs e dall'autore. Tra i primi 24 fori che la cincia ha ispezionato una volta tornata nel locale, 11 erano fori che aveva utilizzato. L'esperimento indica che essa non scopriva i semi spinta da uno stimolo olfattivo. Riuscì a trovare quattro semi *(pallini in colore)*.**

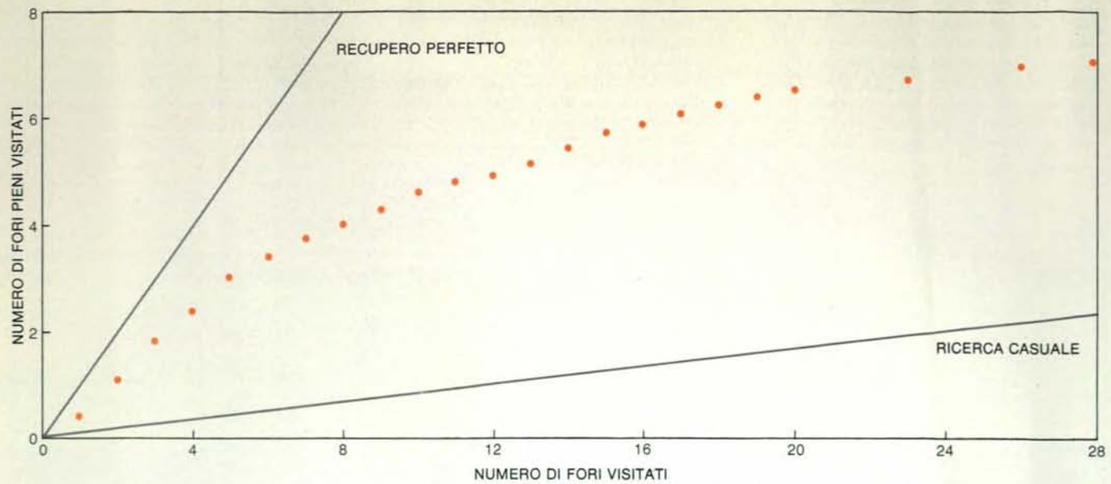





**Nel terzo esperimento, una cincia bigia immagazzinò due lotti di semi. Nello scegliere i fori per il secondo lotto *(cerchi neri con il punto)*, ha evitato quei fori nei quali aveva immagazzinato il primo lotto *(cerchi neri senza punto)*. Questo risultato permette di stabilire che la cincia, sia che immagazzini i semi sia che li vada a recuperare, non visita mai casualmente una particolare serie di fori: essa ricorda dove è già stata.**



**L'ultimo esperimento riguardava il recupero di ambedue i lotti di otto semi da parte della cincia bigia che li aveva immagazzinati. Ispezionando 29 fori, essa ha trovato 10 semi: quattro *(cerchi in colore)* dal primo lotto e sei *(cerchi in colore con il punto)* dal secondo. Tutte le prime sette ispezioni dei fori che la cincia ha compiuto hanno avuto successo. In due casi, l'uccello ha compiuto l'errore di rivisitare un foro.**




**La prestazione della cincia bigia è stata riportata in grafico da Krebs e dall'autore di questo articolo. Se la cincia cercasse a caso otto semi in cento fori, ispezionerebbe perlomeno 12 fori per trovare un seme. Se, invece, la sua memoria fosse perfetta, ispezionerebbe un solo foro per trovare ogni seme. La prestazione della cincia bigia è stata intermedia: in 12 prove ha ispezionato una media di circa 3,7 fori per trovare un seme.**

vicini. Con un solo stimolo da ricordare per volta, i piccioni dell'esperimento di Wilkie e Summers non sembravano riconoscere quale disco fosse stato illuminato solo pochi secondi dopo che l'avevano visto. Per contro, una nocciolaia di Clark ricorda chiaramente per mesi migliaia di siti di immagazzinamento.

Il compito del piccione differiva comunque per vari aspetti da quello della nocciolaia: infatti, il piccione vedeva passivamente lo stimolo che avrebbe dovuto ricordare. Gli stessi stimoli erano invariabilmente a pochi centimetri l'uno dall'altro e differivano solo per la posizione all'interno della matrice. Il cibo con il quale il piccione veniva ricompensato non era dietro lo stimolo. Uno qualsiasi di questi fattori presumibilmente rendeva il compito del piccione più arduo di quello di fronte al quale si trova un uccello che faccia incetta di cibo. Per esempio, indagini di laboratorio effettuate principalmente su ratti mostrano che gli animali possono ricordare per parecchie ore i luoghi che visitano effettivamente.

Può darsi, dunque, che i tipi di stimoli presenti in corrispondenza dei siti dove si accumulano le provviste siano proprio la causa che permette di ricordare quei siti. Secondo questa ipotesi, un uccello che fa incetta di cibo non dovrebbe necessariamente avere una memoria molto superiore a quella di altri animali. L'adozione di un sistema di vita basato sull'incetta di cibo lo avrebbe semplicemente messo nelle condizioni in cui la sua memoria gli diventa particolarmente utile. In che modo lo si potrebbe provare? I confronti, per quanto concerne la memoria, tra specie che immagazzinano il cibo e specie che non lo fanno sono spesso compromessi da differenze nelle motivazioni e nelle possibilità motorie e sensoriali della specie. Ovviamente, i confronti sarebbero più significativi se le specie fossero affini.

Da questo punto di vista, è degno di nota il fatto che, tra i corvidi (cioè le ghiandaie, i corvi e le nocciolaie) e tra le cince alcune specie immagazzinino il cibo mentre altre no. Tra le cince, per esempio, la cincia col ciuffo e la cincia bigia alpestre chiaramente immagazzinano cibo in autunno per nutrirsene poi in inverno, proprio come fa la nocciolaia di Clark. La cincia bigia immagazzina cibo in sovrabbondanza per periodi di tempo relativamente brevi. La cinciarella e la cinciallegra non immagazzinano nulla.

Tra i corvidi, un esempio di varie specie correlate tra loro ma con una diversa tendenza ad accaparrare cibo è stato descritto da Vander Wall, in collaborazione con Russell P. Balda della Northern Arizona University. La nocciolaia di Clark vive nelle stesse aree degli Stati Uniti occidentali dove vivono tre altri membri della famiglia corvidi, e cioè la ghiandaia dal capo azzurro, la cianocitta di Steller e l'afelocoma azzurra. Le quattro specie differiscono in misura notevole per la loro dipendenza dai pinoli immagazzinati e, nel contempo, si diversificano per il modo in cui sono in grado di immagazzinare e di trasportare questi semi.

Di tutte, la nocciolaia di Clark è quella che mostra un maggior grado di dipendenza. Sfruttando le sue provviste di semi, essa è in grado, infatti, di cominciare a riprodursi fin dal mese di febbraio, molto prima della maggior parte degli altri uccelli presenti nella stessa area. Inoltre è anche la più specializzata anatomicamente per raccogliere e fare incetta di semi: oltre a una tasca sublinguale, nella quale pone una grande quantità di semi, possiede un lungo becco affilato con cui riesce ad aprire le pigne ancora verdi prima che possano aprirsi da sole. La nocciolaia di Clark è anche un potente volatore e può trasportare carichi pesanti per grandi distanze fino a siti di immagazzinamento adatti.

Per contro, l'afelocoma azzurra è la meno specializzata delle quattro specie: non ha gli adattamenti anatomici della nocciolaia di Clark e dipende molto meno dalle scorte di semi. Nasconde tutti i semi che mette da parte nell'area dove vive: area che difende come fosse il proprio territorio. La ghiandaia dal capo azzurro e la cianocitta di Steller hanno caratteristiche intermedie sia come anatomia sia come comportamento. La memoria spaziale dei componenti di queste tre specie è forse meno sviluppata di quella della nocciolaia di Clark? La risposta richiede ulteriori ricerche sia in natura sia in laboratorio.

Le ricerche sulla memoria e sull'incetta di cibo che ho descritto sono al limite tra gli studi zoologici e quelli psicologici di comportamento animale. Ricerche di questo stampo si stanno sempre più sviluppando a mano a mano che gli psicologi si convincono che vale la pena di considerare gli studi di laboratorio sull'apprendimento e sulla memoria come mezzi per analizzare in che modo gli animali risolvono i problemi con i quali si trovano a confronto in natura. Nel contempo, gli zoologi si interessano sempre più agli aspetti dell'approvvigionamento del cibo, che richiedono chiaramente una capacità di apprendimento e di memoria. Sforzi congiunti possono dunque associare l'interpretazione zoologica del modo in cui il comportamento si adatta all'ambiente con gli strumenti concettuali dello psicologo e le tecniche sperimentali. Essi sono l'inizio di nuovi e stimolanti sviluppi nello studio della memoria.

# Il futuro dell'universo

*Una previsione per l'espansione dell'universo fino all'anno $10^{100}$: i protoni decadranno, le galassie formeranno buchi neri e i buchi neri «evaporeranno». Se l'universo collassa, può essere ciclico*

di Duane A. Dicus, John R. Letaw, Doris C. Teplitz e Vigdor L. Teplitz

Negli ultimi anni gli sviluppi della fisica delle interazioni ad alta energia delle particelle elementari hanno stimolato una notevole crescita della cosmologia. Il tentativo di descrivere tutte le forze fondamentali della natura come aspetti differenti di una sola forza fondamentale ha avuto un parziale successo nelle «teorie di grande unificazione» delle particelle elementari. Mediante tali teorie si possono presentare descrizioni probabili dei processi fisici dominanti, dalle bassissime temperature a temperature dell'ordine di $10^{32}$ kelvin. Si possono perciò individuare le proprietà della materia a densità di interesse cosmologico, che variano da meno di $10^{-300}$ grammi a più di $10^{100}$ grammi per centimetro cubo. Le condizioni estreme ai confini di queste scale possono prevalere soltanto nei primissimi o negli ultimissimi stadi dell'evoluzione dell'universo.

Alcuni anni fa i cosmologi cominciarono a mostrare come la fisica teorica possa fornire una valida spiegazione della storia dell'universo entro i primi $10^{-35}$ secondi dall'inizio. Più recentemente parecchi fisici e cosmologi, tra i quali noi quattro, hanno estrapolato gli eventi cosmici in un futuro che si estende fino a $10^{100}$ volte l'età stimata dell'universo. A parte il fascino intrinseco di tale impresa, vi sono molti motivi per affrontarla. Dal punto di vista della fisica teorica l'estrapolazione rende possibile un tipo di esperimento ideale: determinando in che modo i diversi effetti previsti da una teoria interagiscono in un dato istante, si può verificare la coerenza globale e la plausibilità di tale teoria. In nessun laboratorio ordinario è probabile che si possano mai raggiungere le temperature e le densità alle quali forniscono previsioni le teorie di grande unificazione e quindi queste teorie vengono sottoposte a verifica in quello che alcuni fisici chiamano l'ultimo laboratorio, cioè l'universo nel suo insieme. Dal punto di vista della cosmologia l'importanza delle nostre estrapolazioni sta nel fatto che le teorie di grande unificazione hanno anche conseguenze che si possono verificare nei grandi laboratori terrestri e si possono perciò confermare le loro previsioni sulle condizioni estreme. Si può pertanto tracciare una sequenza di eventi che potrebbero verificarsi nel lontano futuro con dettagli decisamente superiori a quelli precedentemente possibili.

Il punto di partenza dei nostri calcoli sul futuro remoto è il modello del big bang. Oggi si accetta generalmente che l'evoluzione dell'universo sia stata determinata nei primissimi istanti del big bang. Secondo tale modello, l'universo ebbe inizio con l'esplosione di un'entità primordiale estremamente compatta, fra i 10 e i 20 miliardi di anni fa. Il termine «big bang» e la terminologia relativa, che è quella usata per parlare di un'esplosione, sono abbastanza adatti, perché pare proprio che la materia e l'energia siano state espulse nello spazio da una comune origine. Il termine potrebbe tuttavia anche ingannare, perché non è corretto supporre che l'espansione possa essere osservata dall'esterno. Non c'è, infatti, possibilità di osservazione dall'esterno, perché ciò che sta esplodendo è l'intero universo. Lo stesso spazio sta espandendo nel senso che la distanza tra due galassie qualsiasi sta crescendo a una velocità che dipende dalla loro distanza. Viste dalla nostra galassia, tutte le altre galassie appaiono allontanarsi a una velocità che aumenta di 17 chilometri al secondo per ogni milione di anni luce di distanza. Matematicamente il big bang è descritto dalla teoria della relatività generale di Albert Einstein.

La recessione di galassie lontane viene dedotta dall'osservazione dello spostamento della loro radiazione verso l'estremità rossa dello spettro elettromagnetico. L'espansione e il raffreddamento dell'universo hanno alterato ancor più drammaticamente lo spettro di un altro tipo di radiazione, cioè la radiazione cosmica di fondo a microonde, scoperta nel 1964 da Arno A. Penzias e Robert W. Wilson dei Bell Laboratories. Questa radiazione di fondo sembra provenire con la stessa intensità da tutti i punti del cielo; l'intensità varia con la direzione per meno di una parte su 10 000. Sono state la diffusione e l'annichilazione di particelle elettricamente cariche nei primi stadi del big bang a dare origine a essa. Un tempo, quindi, essa era molto calda, ma da allora si è andata raffreddando, a causa dell'espansione dell'universo, fino a raggiungere una temperatura di circa tre kelvin.

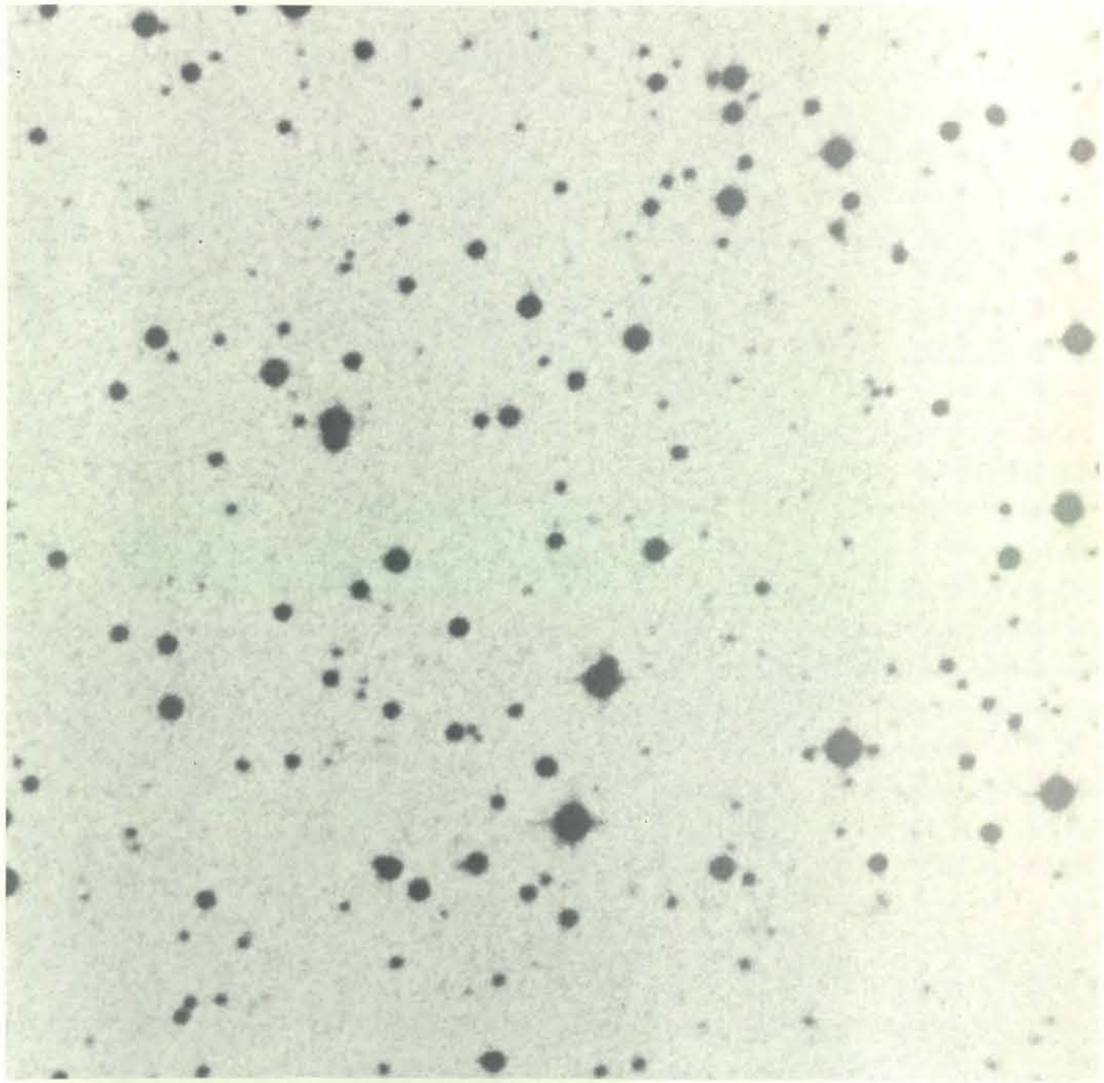
Le reazioni nucleari, alla densità e alla temperatura dell'universo, circa tre minuti dopo l'inizio del big bang furono responsabili della sintesi dell'elio e, in misura molto inferiore, di altri elementi leggeri. L'universo si raffreddò però troppo rapidamente perché potessero formarsi carbonio e altri elementi più pesanti e rimase un'abbondante riserva di idrogeno con la funzione di combustibile nucleare per le stelle. Prima ancora, reazioni avvenute soltanto $10^{-38}$ secondi dopo la nascita dell'universo potrebbero aver generato l'osservato eccesso della materia sull'antimateria, nel qual caso la maggior parte delle teorie di grande unificazione prevedono il decadimento finale di tutta la materia nucleare. Le attuali conoscenze sulle condizioni del primo universo non hanno tuttavia consentito di rispondere alla domanda più incalzante della cosmologia: l'universo continuerà a espandersi per sempre oppure le forze gravitazionali arresteranno la sua espansione e riporteranno tutto lo spazio e il tempo entro i limiti della sfera di fuoco dalla quale è cominciato l'universo?

All'ultima domanda sono stati dedicati molti anni di sforzi osservazionali e di considerazioni teoriche, ma la questione è ancora irrisolta. In linea di principio, il modo più semplice per risolverla è di misurare la quantità di materia in un grande e significativo volume di spazio e di stimare in tal modo la densità della materia su grande scala nell'universo. Se la densità è uguale o superiore a una certa densità critica, che è di circa tre protoni per ogni 1000 litri di spazio, la gravità finirà col vincere l'espansione e l'universo collasserà. Un universo di questo tipo si dice chiuso. Se la densità è inferiore al valore critico, l'universo rimarrà aperto; l'espansione sarà rallentata dalla gravità, ma mai arrestata. Si noti che processi fisici quali la combustione delle stelle o il decadimento di particelle non possono trasformare un universo chiuso in uno aperto.

Stime sul numero di stelle in una galassia tipica e sul numero di galassie per unità di volume forniscono un valore della densità della materia che è solo il 5 per cento di quello critico. Su tale base si potrebbe concludere che l'universo è aperto e che continuerà la sua espansione per sempre. Tuttavia, una stima della massa basata sul conteggio delle stelle comprende soltanto la materia luminosa; se la materia non luminosa costituisce una frazione significativa della massa dell'universo, la stima non è precisa.

Esiste qualche prova dell'esistenza di materia non luminosa nelle galassie. La velocità con cui una stella ruota attorno al centro della sua galassia dovrebbe aumentare con la distanza da tale centro per le stelle che si trovano all'interno della galassia e dovrebbe diminuire con la distanza se la stella si trova alla periferia. Per le stelle esterne, però, la prevista diminuzione della velocità angolare non viene osservata; la distribuzione di velocità osservata porta alla conclusione che alla periferia di una galassia esiste un alone di materia non luminosa la cui massa totale è una o due volte quella della materia galattica luminosa. Quindi la materia non luminosa all'interno delle galassie può fornire un 5-10 per cento della massa necessaria per chiudere l'universo.

Allo stesso modo, i moti delle galassie osservati negli ammassi galattici paiono richiedere materia non luminosa sufficiente a formare il 50 per cento della densità critica dell'universo. La materia non

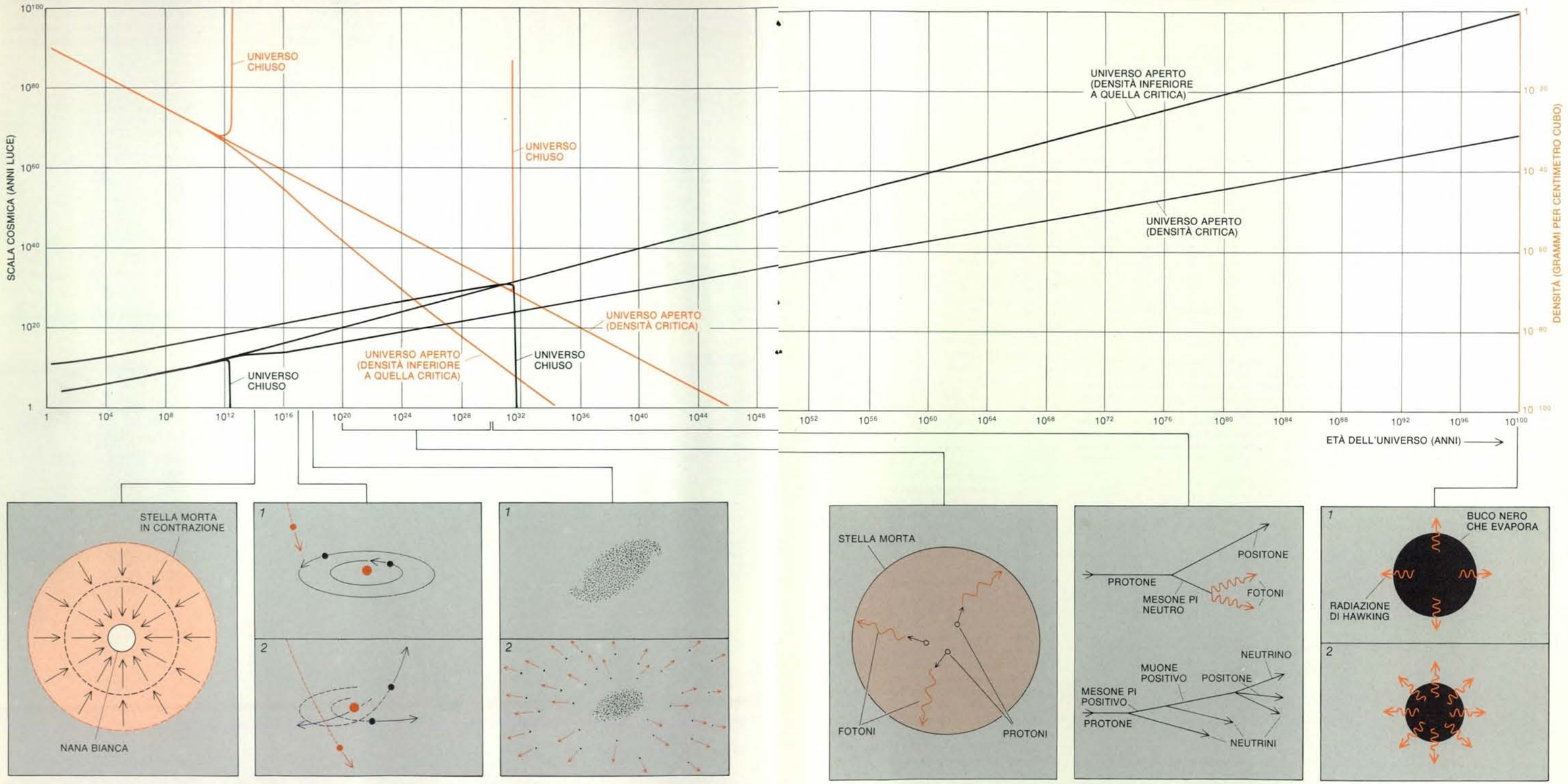


**L'oggetto più lontano conosciuto è un quasar chiamato PKS 2000-330; esso ha uno spostamento verso il rosso di 3,78, cioè si sta allontanando da noi a una velocità pari al 92 per cento della velocità della luce. L'enorme velocità di recessione è probabilmente un effetto dei big bang; se è così, PKS 2000-330 dista più di 16 miliardi di anni luce dalla nostra galassia. La fotografia è una stampa in negativo; il quasar è l'oggetto scuro simile a una stella, al centro. Tali oggetti distanti consentono ai cosmologi di dedurre alcune condizioni dominanti in passato e di estrapolare certi processi nel futuro remoto. La luce che raggiunge oggi la Terra dal quasar deve essere stata emessa subito dopo il big bang.**

luminosa spiegherebbe perché le galassie restano legate in ammassi anziché muoversi in direzioni casuali. La materia oscura potrebbe essere sotto forma di rocce fredde, di stelle combuste, di nubi di gas o persino di buchi neri.

Se la materia ulteriore necessaria per formare il restante 50 per cento della densità critica esiste realmente, potrebbe consistere di atomi di idrogeno ionizzati o di neutrini. L'idrogeno ionizzato dovrebbe essere rilevabile con i telescopi a raggi X in orbita attorno alla Terra. Per quanto riguarda i neutrini, essi dovrebbero essere numerosi quanto i fotoni associati alla radiazione cosmica di fondo; complessivamente, poche centinaia di neutrini per centimetro cubo. Il loro contributo alla densità dell'universo dipende dalla massa del neutrino. Un tempo si riteneva che il neutrino fosse privo di massa, ma gli esperimenti attuali fissano soltanto limiti superiori alle masse del neutrino e non escludono la possibilità che tali masse siano sufficienti a chiudere l'universo. Inoltre, argomentazioni teoriche recenti suggeriscono che il neutrino abbia una qualche massa e alcuni gruppi di ricercatori stanno cercando di rivelarla e di misurarla. Nei prossimi anni tali esperimenti potrebbero rispondere in modo definitivo e affermativo alla domanda se l'universo è chiuso. Del resto, un risultato negativo lascerebbe da risolvere il problema con misure più accurate delle distanze e delle velocità di recessione di galassie lontane.

Anche se l'incertezza attuale sulla densità dell'universo si trova a metà strada tra il valore che caratterizza un universo chiuso e quello di un universo aperto, l'incertezza è limitata a meno di due ordini di grandezza. Il fatto che l'universo non si sia contratto prima di raggiungere la sua età attuale implica che il suo valore di densità non supera di più di 10 volte circa la densità critica. La stima della densità basata sulla materia luminosa indica da



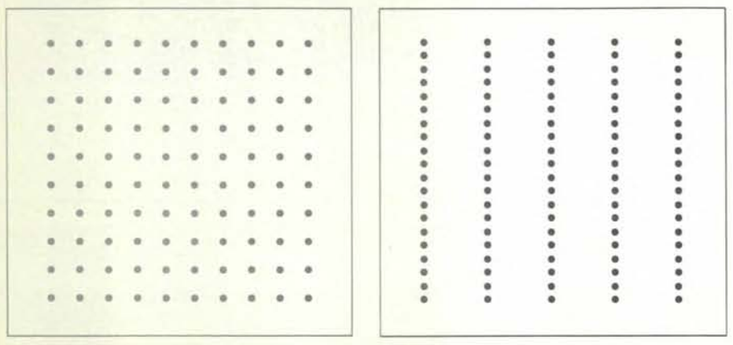

**I sei eventi più importanti del futuro dell'universo sono previsti dalla teoria fisica, purché l'espansione dell'universo iniziata con il big bang continui abbastanza a lungo perché essi si verifichino. Qui la lunghezza di scala (che è circa equivalente al raggio dell'universo) e la densità sono rappresentate graficamente in funzione del tempo per quattro modelli dell'universo: uno nel quale l'universo è aperto, due nel quale esso è chiuso e uno sul confine definito matematicamente tra il modello chiuso e quello completamente aperto. Se l'universo è aperto, la densità della materia e dell'energia è troppo bassa perché la gravità possa vincere l'espansione e quindi l'universo si espanderà per sempre. Se l'universo è chiuso, la gravità finirà coll'arrestare l'espansione e riporterà indietro l'universo in un punto di densità infinita. Non si sa se viviamo in un universo aperto o chiuso. Gli eventi futuri sono schematizzati: le stelle esauriscono il loro combustibile e si contraggono sotto l'azione del loro stesso peso in $10^{14}$anni; perdono i loro pianeti in incontri ravvicinati con le altre stelle in $10^{17}$ anni. Le galassie perdono una grande frazione della loro massa in circa $10^{18}$ anni, a causa di incontri ancor più ravvicinati tra le stelle e a causa della diffusione sia di atomi di idrogeno sia di particelle di polvere; aggregati di materia che acquistano una grande velocità attraverso gli incontri possono vincere le forze gravitazionali all'interno della galassia e la materia restante collassa in un buco nero supermassivo. (Un buco nero è una regione dello spazio dalla quale nulla sfugge se non con i meccanismi descritti dalla teoria dei quanti.) I protoni che decadono generano calore nelle stelle fredde sfuggite dalle galassie. Il riscaldamento diventa significativo dopo $10^{20}$ anni, ed entro l'anno $10^{30}$ circa il 40 per cento della materia dell'universo ha subito tale decadimento. Dopo $10^{100}$ anni i buchi neri supermassivi perdono la loro massa per «evaporazione» quantistica.**

sola che la densità effettiva deve essere almeno un ventesimo di quella critica. Nessuno sa ancora come spiegare quella che i cosmologi trovano essere una sorprendente coincidenza.

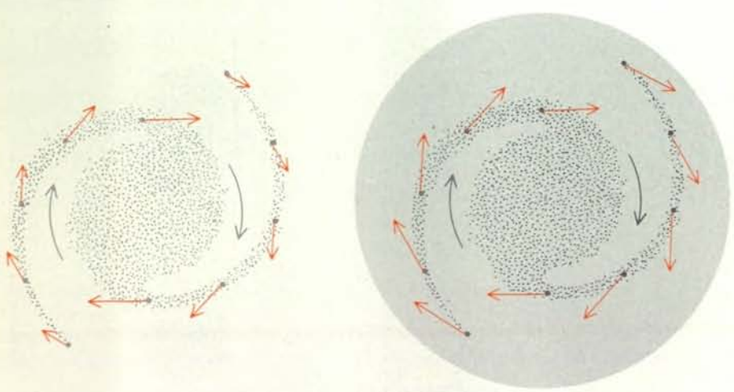
Dal momento che il problema della chiusura non è stato ancora risolto su basi sperimentali o osservazionali, entrambe le possibilità (universo aperto o chiuso) devono essere tenute presenti in qualsiasi speculazione sul futuro lontano. Per cominciare, supponiamo che non venga raggiunta la densità critica e che l'universo sia aperto. Che cosa è probabile che accada? A tale domanda si deve rispondere sia per una struttura su grande scala (cioè per l'evoluzione delle proprietà geometriche dell'universo) sia per la composizione locale (per aggregati di materia le cui dimensioni vanno dai protoni alle galassie).

Secondo le nostre informazioni migliori, la composizione locale dell'universo aperto dovrebbe suscitare sei importanti transizioni. La prima è che entro $10^{14}$ anni dal big bang tutte le stelle avranno esaurito il combustibile. Il principale combustibile nucleare è l'idrogeno, che viene trasformato in elio nel nucleo di una stella per la maggior parte della sua esistenza. Una volta consumata la maggior parte del combustibile idrogeno, la stella si dilata di molte volte le sue dimensioni iniziali e diventa una gigante rossa; in questa fase della sua esistenza l'elio viene solitamente trasformato in carbonio e in elementi più pesanti. Queste reazioni termonucleari sono efficaci transizioni a senso unico: l'idrogeno viene trasformato in elio, l'elio in carbonio e il carbonio in elementi più pesanti in una serie che generalmente termina col ferro. Il nucleo del ferro ha un'energia totale per unità di massa inferiore a quella di qualsiasi altro nucleo, in modo tale che, una volta raggiunto il «limite del ferro», l'energia dell'universo immagazzinata come combustibile nucleare sarà stata interamente liberata.

La velocità alla quale viene consumato il combustibile nucleare di una stella dipende dalla massa della stella: più tale massa è grande, più rapidamente brucia e minore è di conseguenza la sua vita. Per esempio, il Sole consumerà la maggior parte del suo idrogeno entro 10 miliardi di anni circa, successivamente fluttuerà rapidamente di dimensioni, brucerà alcuni degli elementi più pesanti con una rapidità prodigiosa, ma per un tempo relativamente breve, e infine collasserà in una piccola nana bianca che si raffredda lentamente. Le stelle più piccole attraversano tutti questi stadi in un arco di tempo molte volte più lungo della vita del Sole, ma alla fine raggiungeranno anch'esse il limite del ferro. Si noti che, se l'esaurimento del combustibile nucleare è il primo evento di rilievo nel futuro dell'universo aperto, esso si verificherà in un'epoca lontanissima dalla nostra. Le ultime stelle cesseranno di brillare solo quando l'universo avrà un'età 10 000 volte superiore a quella attuale.



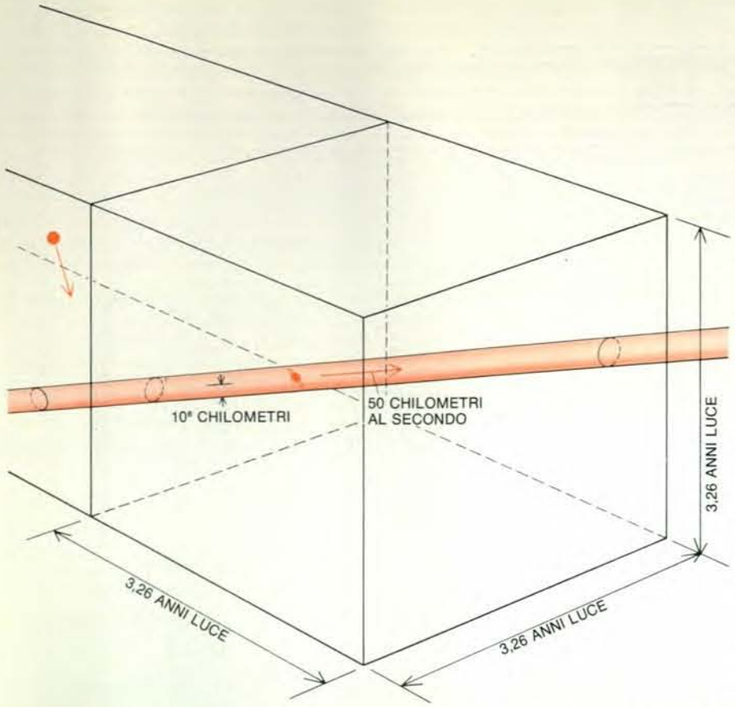
**La distribuzione della materia e dell'energia nell'universo appare sia omogenea sia isotropa con un elevato grado di precisione. In un universo omogeneo numeri uguali di galassie riempiono volumi paragonabili di spazio. In un universo isotropo la densità della materia e dell'energia, oltre a essere omogenea, è la stessa in tutte le direzioni. Le due proprietà non sono equivalenti: un universo omogeneo non è necessariamente anche isotropo. Se la materia è distribuita come nel diagramma schematico a sinistra, l'universo è sia omogeneo sia isotropo; la distribuzione mostrata a destra invece è omogenea ma non isotropa: la distanza tra le galassie lungo la direzione verticale è minore di quella lungo la direzione orizzontale. L'anisotropia in un universo omogeneo in espansione è prodotta da una velocità di espansione che risulta variabile con la direzione.**



**La materia non luminosa può costituire una parte significativa della massa totale dell'universo. Ci sono prove che tale materia costituisce almeno la metà della massa della maggior parte delle galassie. Se le galassie fossero fatte interamente di materia osservabile per via diretta, le stelle alla periferia di una galassia ruoterebbero attorno al centro galattico con una velocità che diminuisce con la distanza dal centro *(a sinistra)*. Invece la velocità osservata per le stelle periferiche non diminuisce con la distanza. Questa discrepanza si può spiegare supponendo che vi sia materia non luminosa distribuita nella galassia e in un alone attorno a essa. Perciò, le stelle che appaiono alla periferia si troverebbero in realtà in mezzo alla materia galattica e la distribuzione delle loro velocità con la distanza dal centro galattico riprodurrebbe più da vicino la distribuzione osservata *(a destra)*. Le lunghezze delle frecce negli schemi sono proporzionali alle velocità delle stelle.**

Il secondo evento importante è che tutte le stelle perderanno i loro pianeti. Se una stella con un pianeta viene avvicinata da un'altra stella entro il raggio dell'orbita planetaria, l'orbita verrà sensibilmente deformata dal campo gravitazionale della stella che passa e il pianeta può essere disperso nello spazio. Il tempo medio di attesa di un tale incontro dipende dalla densità delle stelle in una data regione, dall'area delle orbite planetarie e dalla velocità relativa delle stelle. La densità delle stelle si può esprimere in funzione del volume nel quale esiste la probabilità di trovare almeno una stella. Una stella che si muove nello spazio con un pianeta traccia un volume cilindrico le cui dimensioni dipendono dall'area dell'orbita del pianeta e dalla velocità della stella. L'intervallo medio tra due incontri di stelle è uguale al tempo necessario perché il volume del cilindro diventi uguale al volume nel quale ci si può aspettare una stella *(si veda lo schema a pagina 76)*.

La densità delle stelle in una galassia è di una stella per 35 anni luce cubi. Freeman J. Dyson dell'Institute for Advanced Study di Princeton avanza l'ipotesi che un valore ragionevole del raggio di un'orbita planetaria è di circa 100 milioni di chilometri e fa notare che una tipica stella trascina i suoi pianeti nello spazio a una ve-




**La frequenza degli incontri tra una stella e un sistema planetario può essere stimata dal numero di stelle per unità di volume, dalle dimensioni del sistema planetario e dalla sua velocità rispetto alle stelle. Gli incontri perturbano un'orbita planetaria ogniqualvolta la stella interseca il volume di spazio occupato dall'orbita del pianeta. In media un tale incontro avverrà nello stesso tempo necessario perché il sistema planetario tracci un cilindro nello spazio con volume uguale al cubo all'interno del quale ci si aspetta di trovare una stella. Esiste una stella nella galassia per ogni 35 anni luce cubi. Ogni sistema planetario si muove a circa 50 chilometri al secondo e le orbite planetarie esterne tracciano un cilindro il cui raggio è di circa 100 milioni di chilometri. Con tali presupposti si può subito calcolare che un incontro si verifica circa ogni $10^{15}$ anni. Dopo circa 100 incontri, che si prevedono in $10^{17}$ anni, le orbite dei pianeti saranno state tanto sconvolte che i pianeti saranno sfuggiti dalla stella madre e si saranno dispersi nello spazio interstellare. La scala delle stelle e del sistema planetario rispetto alla loro densità nello spazio è stata esagerata.**

locità di circa 50 chilometri al secondo. Il volume del cilindro tracciato dal sistema planetario in movimento sarà perciò di 35 anni luce cubi dopo circa $10^{15}$ anni e in tale intervallo è quindi probabile un incontro distruttivo. Si può supporre con certezza che dopo 100 di tali incontri tutti i pianeti di una stella siano stati espulsi dall'orbita, di modo che in circa 100 volte $10^{15}$ anni, cioè $10^{17}$ anni, tutte le stelle avranno perso i loro pianeti.

La terza transizione prevista è il risultato di incontri stellari ancor più ravvicinati; i loro effetti sono evidenti su scala galattica. Quando due stelle si avvicinano tra loro, l'interazione gravitazionale può trasferire energia cinetica da una stella all'altra. Se l'incontro è abbastanza ravvicinato, una stella può acquistare tanta energia da raggiungere la velocità necessaria per sfuggire dalla galassia. Dato che nell'interazione si conserva l'energia, la seconda stella deve perdere energia cinetica e come risultato si lega più strettamente al nucleo della galassia.

Il processo può essere chiamato evaporazione galattica, poiché le interazioni stellari riproducono su grandissima scala le interazioni delle molecole che evaporano dalla superficie di un liquido. Uno scambio di energia di questo tipo può anche far sfuggire dalla galassia una frazione apprezzabile del gas interstellare. Dopo l'evaporazione di circa il 90 per cento della massa della galassia, il campo gravitazionale trascinerà le stelle e la polvere rimanenti in un nucleo sempre più denso. È probabile che le galassie nella loro forma attuale comprendano un buco nero centrale supermassivo, una regione di spazio dalla quale è impossibile sfuggire (tranne che con i meccanismi descritti dalla teoria dei quanti). Se oggi questo buco nero non esiste, il nucleo galattico crescerà probabilmente ancora, diventando talmente denso che la gravitazione vincerà la resistenza offerta dalla pressione del gas e il nucleo collasserà comunque per formare un buco nero supermassivo. Un calcolo simile a quello che abbiamo indicato per la perdita dei pianeti mostra che l'evaporazione delle stelle e il collasso delle galassie dovrebbero completarsi dopo circa $10^{18}$ anni.

La quarta e la quinta transizione che immaginiamo per l'universo aperto sono effetti cosmologici ritardati previsti dalla maggior parte dei modelli di grande unificazione della fisica delle particelle, ma tali effetti non diventano importanti fino a quando l'universo non ha un'età 100 volte maggiore di quella in cui le galassie collassano. L'obiettivo delle teorie di grande unificazione è di presentare una spiegazione che comprenda le forze forte, debole ed elettromagnetica tra le particelle elementari. Alle energie relativamente basse raggiungibili nei laboratori terrestri le tre forze appaiono del tutto diverse; ciononostante, le suddette teorie le descrivono come manifestazioni di una sola interazione, la cui unità si può percepire all'energia corrispondente a una temperatura dell'ordine di $10^{27}$ kelvin. A tale straordinaria temperatura particelle quali i quark, che «sentono» la forza forte, si possono trasformare in particelle quali l'elettrone e il positone, che a energie minori sentono soltanto le forze debole ed elettromagnetica. A temperature inferiori tali cambiamenti di identità diventano altamente improbabili, ma possono ancora verificarsi occasionalmente. Il protone potrebbe decadere se i suoi tre quark costituenti subissero una trasformazione.

Secondo la maggior parte delle teorie di grande unificazione, tutti i protoni dovrebbero decadere dopo un periodo di $10^{30}$ - $10^{32}$ anni; un tasso di decadimento di almeno un protone all'anno dovrebbe perciò venire registrato in qualsiasi aggregato che contenga $10^{32}$ protoni, quale una massa d'acqua che pesa circa 160 tonnellate. Vi sono oggi 13 esperimenti, o in atto o in studio, che tentano di controllare grandi masse d'acqua, di ferro o di altri materiali per evidenziare il decadimento del protone. Segnali di decadimento sono stati probabilmente identificati da 150 tonnellate di ferro controllate a 2300 metri di profondità nel campo aurifero di Kolar presso Bangalore nell'India meridionale. Un unico segnale di decadimento è stato registrato dai rivelatori installati nel tunnel del Monte Bianco tra la Francia e l'Italia. È troppo presto per giudicare se tali segnali resisteranno alle analisi.

Se il protone è suscettibile di decadimento, il processo avrà un effetto significativo sulle stelle che non sono state catturate da buchi neri galattici. Tali stelle sono le sole che sfuggono per evaporazione e il decadimento dei protoni e dei neutroni in esse le manterrà notevolmente più calde del mezzo interstellare circostante. Ammettendo una vita media del protone di $10^{30}$ anni, il tasso di decadimento di una tipica stella con le stesse dimensioni del Sole è di circa $10^{27}$ all'anno. Ogni decadimento di un protone dà origine a uno sciame di elettroni, positoni, neutrini e fotoni energetici. Tutte le particelle figlie, tranne i neutrini, vengono assorbite dalla stella e questa energia assorbita mantiene calda la stella.



Si può determinare l'esatta temperatura della stella durante l'era di decadimento dei protoni uguagliando il tasso di irraggiamento di energia dalla stella al tasso di liberazione di energia termica da parte del decadimento. In questo stato di equilibrio, la temperatura dipende dalla massa della stella, dalla superficie dalla quale può essere irraggiato il calore e dall'energia di riposo e dalla vita media del protone. La temperatura di equilibrio è di 100 kelvin per le stelle morte con maggior massa (che, paradossalmente, sono le più piccole) e di circa tre kelvin per le stelle più grandi e con minor massa.

Le stelle si raffredderanno alla loro temperatura di equilibrio in circa $10^{20}$ anni e da quel momento in poi la loro temperatura si manterrà pressoché costante finché verso l'anno $10^{30}$ sarà decaduta la maggior parte dei protoni. Il bagliore stellare è freddo, ma non in confronto alla temperatura della radiazione di fondo residua del big bang. La temperatura di fondo dipende dai particolari della espansione dell'universo aperto. Se la densità dell'universo è minore di quella critica, la temperatura della radiazione di fondo sarà scesa a $10^{-20}$ kelvin entro l'anno $10^{30}$. Se invece la densità è esattamente uguale a quella critica, l'universo si sarà espanso più lentamente e la temperatura della radiazione di fondo sarà scesa a $10^{-13}$ kelvin. Cioè, la temperatura sarà da 13 a 20 ordini di grandezza inferiore alla temperatura delle stelle morte.

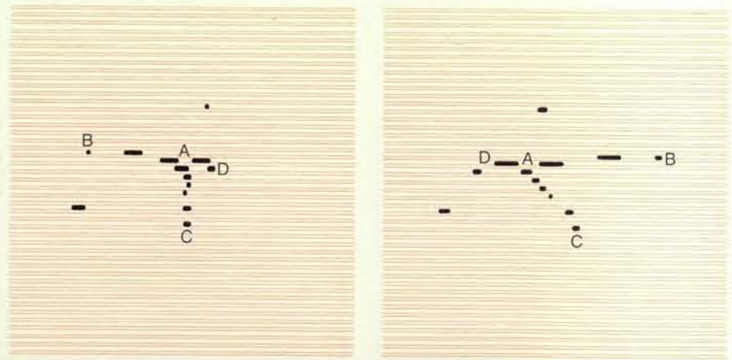
Il decadimento del protone fa anche cambiare la costituzione dei gas interstellari che evaporano prima che le galassie collassino. All'interno di una stella un positone liberato dal decadimento di un protone incontra subito un elettrone e le due particelle si annichilano. L'annichilazione produce altri fotoni e riscalda la stella. Nello spazio intergalattico la densità della materia è così bassa (e in continua diminuzione a causa dell'espansione dell'universo) che è estremamente improbabile che positoni ed elettroni entrino in collisione. In effetti, entro l'anno $10^{30}$ un universo aperto con una densità inferiore a quella critica si sarà espanso a più di $10^{20}$ volte le sue dimensioni attuali e la distanza media tra un elettrone e un positone nel mezzo interstellare sarà dello stesso ordine di grandezza delle dimensioni della nostra galassia. (Se la densità dell'universo è uguale a quella critica, l'universo si sarà espanso di un fattore $10^{13}$.) Il mezzo interstellare diventa perciò un gas estremamente rarefatto formato da circa l,1 per cento degli elettroni di oggi e dai positoni creati dal decadimento di circa l'1 per cento dei protoni di oggi.

Gli eventi associati al decadimento del protone si saranno rivelati quando l'esistenza dell'universo avrà raggiunto i $10^{32}$ anni, un arco di tempo 100 volte superiore alla vita media del protone. Ciò che rimane dell'universo a questo punto sono il gas rarefatto di elettroni e positoni, i fotoni e i neutrini emessi in varie epoche precedenti e i buchi neri supermassivi. I fotoni e i neutrini sono residui del big bang, degli anni in cui le stelle splendevano, del decadimento di protoni e neutroni attraverso la storia e del decadimento finale di stelle morte. I fotoni e i neutrini perdono energia e sia essi sia gli altri costituenti diventano più rarefatti col continuare della espansione. Per altri aspetti l'universo a densità inferiore a quella critica resta quiescente fino a circa l'anno $10^{100}$, un periodo $10^{68}$ volte più lungo di tutti i processi che abbiamo descritto finora.

Il sesto e ultimo evento di rilievo nel futuro dell'universo aperto è il decadimento dei buchi neri. Nell'interpretazione più semplice della teoria gravitazionale di Einstein nulla può sfuggire da un buco nero. Esiste un confine, chiamato orizzonte degli eventi, lungo il quale la velocità necessaria per sfuggire è uguale alla velocità della luce; nessuna particella all'interno dell'orizzonte degli eventi può muoversi tanto rapidamente da attraversarlo. Ciononostante, nel 1974, S. W. Hawking dell'Università di Cambridge dimostrò che un fenomeno quantomeccanico comporta che un buco nero possa emettere tutta l'energia associata alla sua massa e quindi scomparire.

Il fenomeno quantomeccanico è una conseguenza del principio di indeterminazione di Werner Heisenberg, secondo il quale la quantità di moto o la posizione di una qualsiasi particella ha soltanto una certa probabilità di cadere all'interno di un dato intervallo di valori. Più esattamente, il prodotto dell'indeterminazione della quantità di moto della particella per l'indeterminazione della sua posizione non è inferiore a una costante numerica. In fisica classica (cioè, nella teoria fisica che non tiene conto dei fenomeni quantistici) una particella può attraversare una barriera energetica solo se le si conferisce energia sufficiente a superare tale barriera. L'orizzonte degli eventi è in fisica classica una barriera energetica assoluta: una particella non potrebbe acquistare energia sufficiente a superarlo. Invece, a causa del principio di indeterminazione, una particella che si trova inizialmente in una regione può trovarsi successivamente in un'altra regione, anche se la sua energia classica è molto inferiore all'altezza della barriera energetica tra le due regioni. Si dice che una particella, che attraversa una barriera energetica in questo modo, senza acquistare l'energia necessaria a superarla, ha attraversato la barriera per effetto tunnel. Hawking sottolineò che, dal momento che le particelle possono attraversare l'orizzonte degli eventi per effetto tunnel, da un buco nero possono essere espulse massa ed energia (si veda l'articolo *La meccanica quantistica dei buchi neri* di S. W. Hawking, in «Le Scienze», n. 105, maggio 1977).

Hawking dimostrò che il tasso di emissione di energia da un buco nero è inversamente proporzionale al quadrato della sua massa. Il tasso è inizialmente basso ma, al diminuire della massa, la perdita di energia aumenta. Ne consegue che tutti i buchi neri devono alla fine scomparire o evaporare; per i buchi neri supermassivi, resti di galassie collassate, il tempo di evaporazione è di circa $10^{100}$ anni. La maggior parte dei prodotti di decadimento sono fotoni; l'emissione diventa sempre più energetica negli ulti-




**Il decadimento del protone è un possibile evento che dovrebbe avere conseguenze importanti per il futuro dell'universo. Esso dovrebbe osservarsi sperimentalmente se non è troppo raro. La maggior parte delle attuali «teorie di grande unificazione» delle interazioni tra particelle elementari prevede che la durata della vita del protone sia compresa tra $10^{30}$ e $10^{32}$ anni. Ben 13 esperimenti in corso o in fase di progettazione stanno cercando di identificare il decadimento. Sono all'esame alcuni eventi, che potrebbero essere di decadimento, registrati in due esperimenti. Alcuni di questi eventi sono stati rivelati da ricercatori che operano in una miniera nei campi auriferi di Kolar nei pressi di Bangalore nell'India meridionale; un altro evento è stato visto da ricercatori che operano presso l'Organizzazione europea per le ricerche nucleari (CERN), che lavorano con un rivelatore installato nel tunnel del Monte Bianco tra la Francia e l'Italia. L'evento del Monte Bianco è illustrato in due vedute schematiche, che risultano una perpendicolare all'altra. I piani paralleli sono lastre di ferro controllate per eventuali decadimenti e le macchie nere rappresentano i prodotti del decadimento. Se si riesce a verificare il decadimento del protone, si può fissare una pietra miliare dell'evoluzione dell'universo attorno all'anno $10^{30}$. A quel tempo sarà decaduta tutta la materia a lunga vita, fatta eccezione per gli elettroni, i positoni, i fotoni, i neutrini e i buchi neri.**

mi stadi dell'evaporazione. Perciò, dopo $10^{100}$ anni, l'universo è costituito da un gas estremamente diffuso di elettroni, positoni e neutrini, da fotoni di bassa energia emessi molto tempo prima del decadimento dei buchi neri e da numerose sfere in espansione di fotoni di alta energia emessi dall'evaporazione di buchi neri.

I modelli matematici della struttura su grande scala di un universo aperto indicano che l'universo nel quale viviamo si trova in una condizione molto particolare. Per descrivere matematicamente l'universo si devono precisare sei grandezze: tre danno la velocità di espansione lungo le tre direzioni dello spazio e tre danno il tasso di variazione della velocità di espansione (la sua accelerazione o decelerazione) lungo le tre direzioni. Se le sei grandezze si definiscono in un certo istante e se si procede in questo istante alla distribuzione di energia, si possono determinare tali valori e la distribuzione energetica in tutti gli istanti successivi.

Nel 1973 Hawking e C. B. Collins, pure di Cambridge, dimostrarono, che per quasi tutti i valori possibili di queste grandezze in una fase primitiva durante il big bang, l'universo deve diventare sempre più anisotropo mentre si espande. La distanza media delle galassie lungo un asse spaziale dovrebbe diventare sempre più diversa dalla distanza media lungo gli altri assi a causa delle differenze nelle velocità di espansione. Il risultato fa pensare che l'attuale universo sia estremamente improbabile. Per esempio, come abbiamo fatto notare prima, il fondo cosmico a microonde è invariante con la direzione entro una parte su 10 000. Su scala cosmica, anche la materia dell'universo è distribuita regolarmente. Un altro problema irrisolto della cosmologia è quello di capire perché l'universo sia in questo modo quasi isotropo.

Un modo di affrontare il problema è di scegliere valori iniziali delle sei gradezze in modo tale che la densità di energia della materia e della radiazione sia esattamente uguale alla densità critica. Se nella sua storia un siffatto universo è dominato dalla materia anziché dalla radiazione, qualsiasi anisotropia iniziale non potrà accrescersi. Il decadimento del protone, però, può portare a un universo dominato dalla radiazione.

John D. Barrow e Frank J. Tipler dell'Università della California a Berkeley hanno dimostrato che in un universo alla densità critica gli elettroni e i positoni che rimangono dopo il decadimento del protone cominceranno a formare coppie legate dopo più di $10^{70}$ anni. Ogni coppia legata è in realtà un atomo, ma l'elettrone e il positone orbitano ciascuno in uno stato così altamente eccitato che la regione di spazio occupata dalla coppia è più estesa dell'universo attualmente osservabile. Lo spazio occupato diventa tanto più grande quanto più tardi si forma una coppia legata.

Col passar del tempo l'elettrone e il positone si muovono a spirale verso l'interno e finiscono con l'annichilarsi, dando origine a fotoni di alta energia. La lunghezza d'onda dei fotoni viene spostata verso l'estremità rossa dello spettro mentre l'universo continua a espandersi. Nel modello descritto da Barrow e Tipler esso finisce con l'essere dominato dalla radiazione e l'anisotropia aumenta. Tuttavia, in una ricerca successiva, Don N. Page e M. Randall McKee della Pennsylvania State University dimostrarono che in certe condizioni le coppie elettrone-positone possono continuare a dominare l'energia dell'universo anche se il loro numero è in diminuzione. Lo spostamento verso il rosso della radiazione di fotoni fa diminuire la sua energia e il lungo tempo progressivamente crescente per l'annichilazione di coppie legate, di formazione ritardata, provoca una diminuzione nel tasso di emissione dei fotoni. Quindi può darsi che l'anisotropia caratteristica di un universo che è dominato dalla radiazione non sia più in grado di crescere.

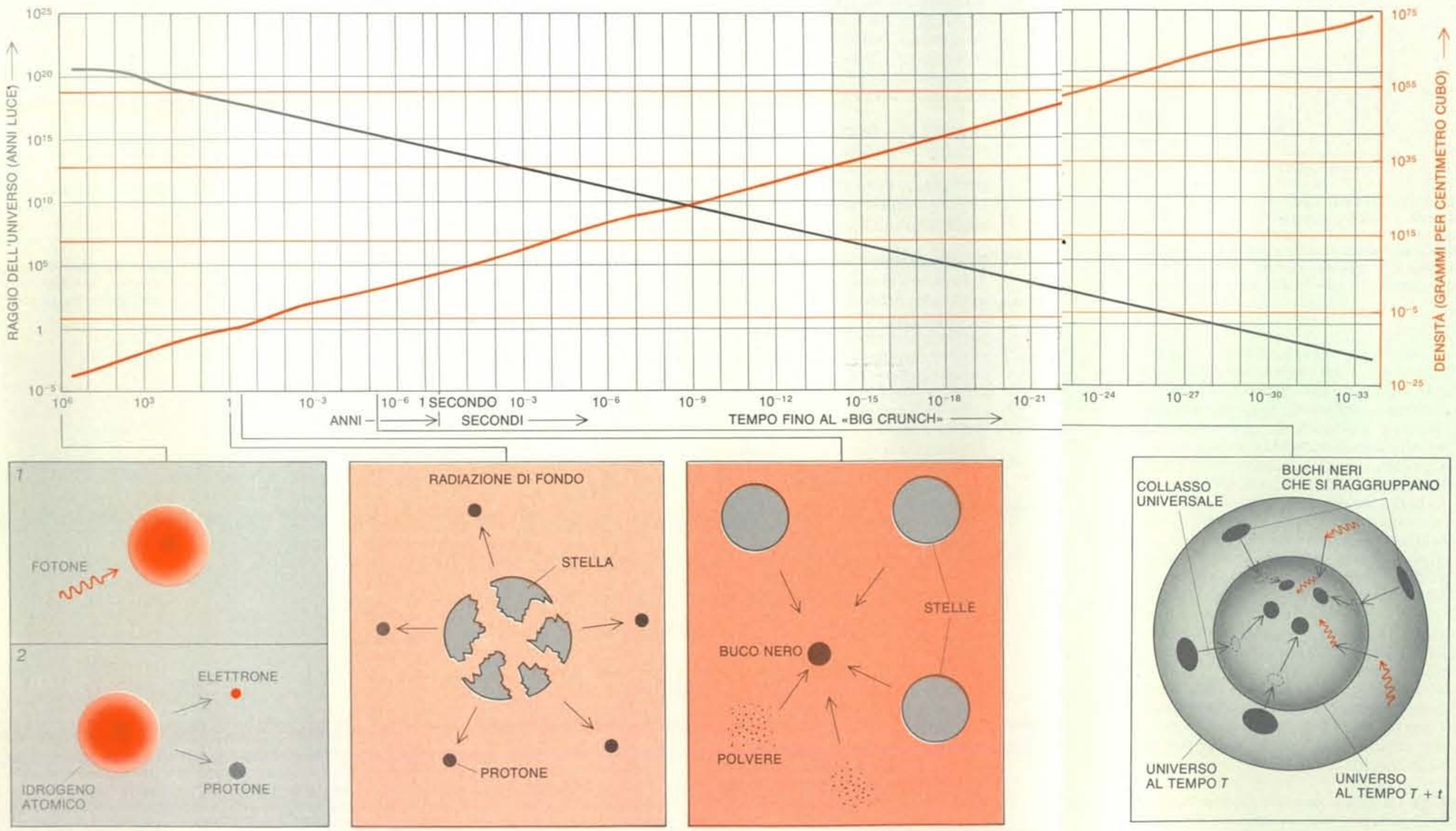
Tutte le speculazioni precedenti riguardano un universo aperto. Analizzeremo ora il futuro dell'universo nell'ipotesi che vi sia sufficiente materia non luminosa perché la gravità arresti l'espansione e origini una contrazione. Quanto più la densità media è vicina a quella critica, tanto più a lungo durerà l'universo chiuso. Non conosciamo però alcun motivo per il quale la densità media debba essere abbastanza vicina alla densità critica da permettere all'universo di sopravvivere fino a che non decadrà la maggior parte dei protoni. È perciò probabile che l'universo, nella sua massima espansione, sia formato da stelle morte, da buchi neri supermassivi, resti di galassie collassate, e da fotoni e neutrini dotati di bassa energia, proprio come accadrebbe se l'universo fosse aperto.

Un aspetto curioso di un universo chiuso è che, anche se l'energia si conserva localmente, la massa o l'energia totale dell'universo non si conserva. Per una data dimensione dell'universo l'energia totale durante la contrazione è maggiore che durante l'espansione. Si consideri un fotone emesso dal Sole verso lo spazio intergalattico.

Il principio di conservazione viene soddisfatto durante l'emissione perché l'energia trasportata dal fotone è esattamente bilanciata da una piccolissima diminuzione della massa del Sole. Mentre l'universo si espande, la lunghezza d'onda del fotone aumenta proporzionalmente, con il risultato che la sua energia diminuisce. Quando l'universo si contrae, diminuisce anche la lunghezza d'onda del fotone e la sua energia per contro aumenta.

Alla fine la lunghezza d'onda del fotone diventa inferiore a quella che il fotone possiede all'istante dell'emissione, e quindi il fotone ha acquistato energia senza alcuna corrispondente perdita di massa o di energia in altre regioni. L'universo si mantiene perciò più caldo durante la contrazione che non durante l'espansione. I fotoni che danno il maggior contributo all'energia supplementare sono quelli emessi quando l'universo si trova alle sue massime dimensioni o quasi, dato che tali fotoni subiranno la massima contrazione.

Gli eventi più importanti durante la fase di un universo chiuso seguono la




**Le fasi finali del collasso di un universo chiuso ripercorrerebbero all'incirca le stesse fasi della sua espansione se non vi fossero gli effetti dei buchi neri. Circa 20 miliardi di anni prima del «big crunch», il collasso gravitazionale completo di un universo chiuso, l'universo si contrarrà in un volume nel quale la sua densità di energia sarà la stessa di oggi. Mentre la contrazione continuerà, i fotoni diventeranno più energetici e l'universo si riscalderà; circa un milione di anni prima del «big crunch» i fotoni dissocieranno gli atomi di idrogeno interstellari nei loro elettroni e protoni. Un anno prima, invece, la temperatura all'esterno delle stelle diventerà maggiore di quella interna e le stelle cominceranno a spaccarsi. Circa nello stesso tempo i buchi neri supermassivi cominceranno a inghiottire il materiale stellare restante, come pure altra materia e radiazione. Circa tre minuti prima, infine, i buchi neri supermassivi cominceranno a raggrupparsi e questo loro raggruppamento è rappresentato dalla contrazione della superficie di una sfera, che corrisponde alla contrazione dello spazio. La variazione di densità e di raggio dell'universo è rappresentata in funzione del tempo; essa è stata calcolata numericamente applicando le equazioni di Einstein e altri principi descritti nel testo.**


## TECNOLOGIA

LE SCIENZE *edizione italiana di* SCIENTIFIC AMERICAN *ha pubblicato su questo argomento numerosi articoli tra cui:*

**TECNOLOGIA E PAESI IN VIA DI SVILUPPO**
di G. Myrdal (n. 79)

**AUTOMATISMI**
di J. S. Albus e J. M. Evans, Jr. (n. 94)

**IL SISTEMA MONDIALE DI TELECOMUNICAZIONI VIA SATELLITE**
di B. I. Edelson (n. 106)

**GLI IMPIEGHI DELLA RADIAZIONE DI SINCROTRONE**
di E. L. Rowe e J. H. Weaver (n. 110)

**COMUNICAZIONI SU ONDA LUMINOSA**
di W. S. Boyle (n. 112)

**COME CONSERVARE L'ENERGIA**
di G. B. Zorzoli (n. 115)

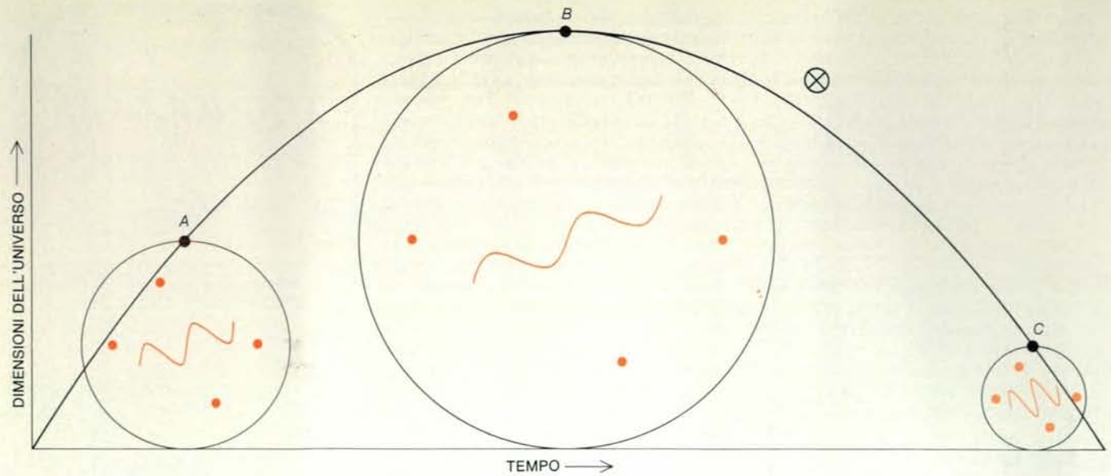
**MONTAGGIO GUIDATO DAL CALCOLATORE**
di J. L. Nevins e D. E. Whitney (n. 116)

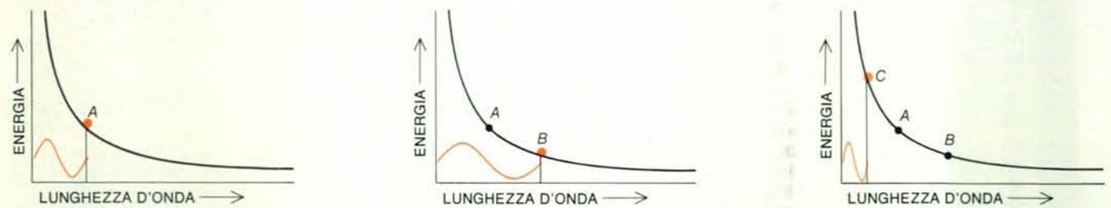
**LA CONVERSIONE BIOLOGICA DELL'ENERGIA SOLARE**
di I. F. Quercia e P. Quercia (n. 119)

**PROGETTI ALTERNATIVI PER IL MOTORE D'AUTOMOBILE**
di D. G. Wilson (n. 121)

**APPLICAZIONI INDUSTRIALI DELLE MEMBRANE SINTETICHE**
di H. P. Gregor e C. D. Gregor (n. 121)

**LA CENTRIFUGAZIONE GASSOSA**
di Donald R. Olander (n. 122)




**Le energie delle particelle libere non si conservano in un universo chiuso. A causa dell'espansione e della contrazione cosmica di un tale universo, i fenomeni che dipendono dalla distanza variano nel tempo, allo stesso modo in cui la distanza tra due punti sulla superficie di un pallone aumenta o diminuisce quando il pallone si gonfia o si sgonfia. Così la lunghezza d'onda di un fotone aumenta durante l'espansione e diminuisce durante la contrazione. L'energia del fotone è proporzionale all'inverso della lunghezza d'onda e quindi l'energia del fotone emesso diminuisce durante l'espansione e aumenta durante la contrazione. Quando la lunghezza d'onda del fotone è minore di quella all'istante della sua emissione, l'universo nell' insieme acquista energia. L'energia addizionale potrebbe originare una nuova, successiva espansione.**

stessa successione degli eventi in un universo aperto. (Non esiste, naturalmente, alcun collasso per un universo aperto.) Numerosi ricercatori tra i quali Martin J. Rees di Cambridge, hanno studiato il collasso. Mentre i fotoni acquistano energia durante la contrazione, essi riscaldano le stelle morte, provocando in tal modo l'inizio di una loro rapida combustione, esplosione o evaporazione.

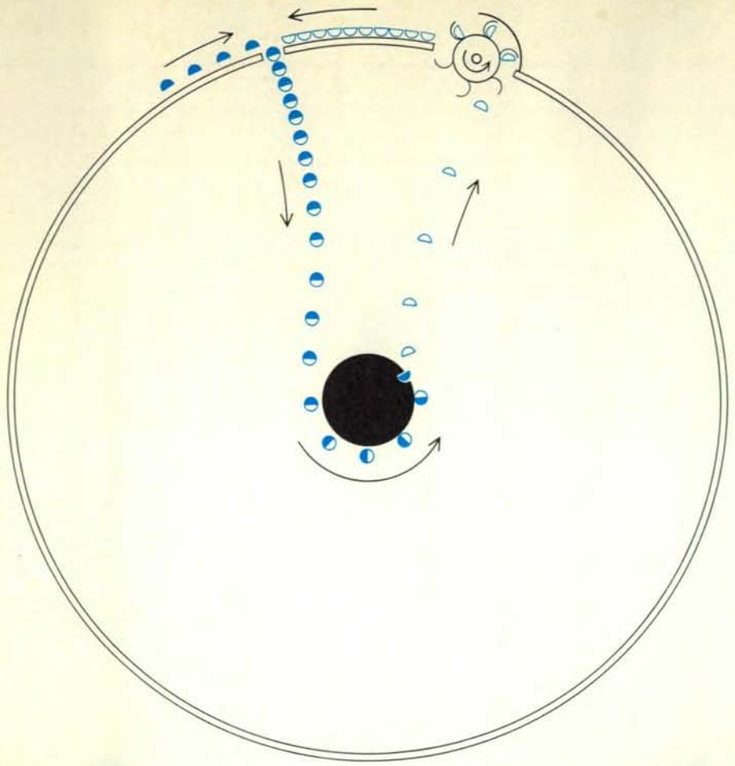
Il risultante «brodo» di particelle continuerebbe a ripetere le varie fasi dell'espansione dell'universo se non fosse per gli effetti dei buchi neri. All'aumentare della densità, i buchi neri ingoiano materia e si fondono quando si scontrano. Si può calcolare che in un universo con un buco nero supermassivo per galassia le stelle morte vengano inghiottite dai buchi neri subito dopo che si spezzano e da esse comincia a evaporare materia. Alla fine tutti i buchi neri si riuniscono in un solo grande buco nero che abbia la stessa ampiezza dell'universo.

La fisica teorica nel suo attuale stato di sviluppo non può descrivere completamente il collasso di un buco nero; non si sa ancora come estrapolare le equazioni che governano la struttura su grande scala dell'universo fino a riportarla a uno stato di densità infinita. È però possibile che, prima che la densità diventi infinita, un meccanismo sconosciuto induca una «inversione di rotta» e faccia sì che l'universo ricominci ancora a espandersi.

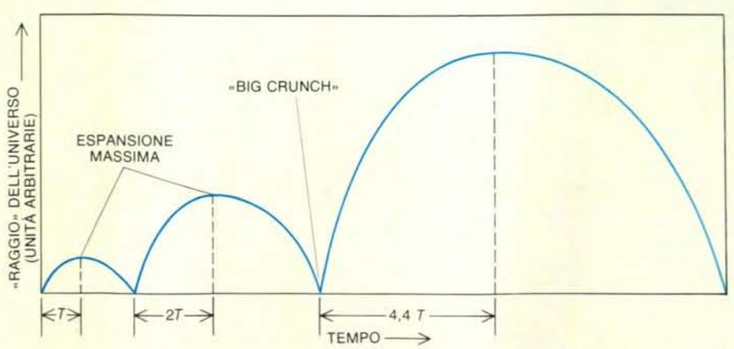
Se l'universo può invertire rotta, esso potrebbe rimanere chiuso e quindi essere ciclico. L'energia acquistata dai fotoni durante ogni periodo di contrazione potrebbe conservarsi durante l'inversione; con ogni successivo ciclo l'universo diventerebbe perciò più grande, per una data temperatura, di quanto non fosse nel ciclo precedente e impiegherebbe anche più tempo per raggiungere le sue massime dimensioni. Robert H. Dicke e P. J. E. Peebles della Princeton University assieme a noi quattro hanno calcolato che, se l'universo è ciclico, il prossimo ciclo dovrebbe espandersi per una durata doppia della fase di espansione del ciclo attuale. Per i cicli precedenti il fattore di espansione sarebbe minore; abbiamo calcolato che l'universo attuale si trova al massimo a 100 inversioni dal ciclo che è durato appena da produrre una sola generazione di stelle. La maggiore difficoltà del modello per le inversioni è quella di capire come un universo straordinariamente disomogeneo e localmente anisotropo, costituito da buchi neri che si stanno gradatamente raggruppando, possa diventare regolare; per tale motivo non si riesce a capire come possa avvenire un'inversione.

Se l'universo è ciclico, e se le reazioni mediate dalle alte temperature esistenti all'inizio di ogni ciclo danno origine all'eccesso di particelle nucleari sulle antiparticelle, oggi osservato, si può trarre un'importante conclusione. L'evoluzione dell'universo dall'inizio di ogni inversione, fino alla formazione e alla dissoluzione delle galassie, seguirà lo stesso corso di eventi durante ogni ciclo. Tuttavia, all'allungarsi dei cicli, i fenomeni che abbiamo descritto per i successivi stadi dell'universo aperto diventeranno importanti. Per esempio, nel primo ciclo in cui la fase di espansione dura più di un decimiliardesimo della vita media del protone (cioè, più di circa $10^{20}$ anni) l'energia acquistata durante la fase di contrazione sarà dominata dai fotoni emessi nel corso del decadimento del protone. I cicli successivi possono perciò diventare più lunghi di un fattore pari a 1000 anziché di un fattore soltanto di 2 circa. Dagli effetti di raggruppamento di buchi neri potrebbero risultare fattori di espansione perfino maggiori, dell'ordine di $10^{12}$.

Da un punto di vista umano forse la domanda più importante che ci si potrebbe porre sull'universo riguarda il futuro della vita e dell'intelligenza. In un universo sempre più ostile alla vita, può continuare indefinitamente l'abitudine di pensare come accade sulla Terra? Numerosi cosmologi, tra i quali si ricordano Dyson e Steven C. Frautschi del Califor-



**La crisi energetica cosmica potrebbe essere molto ritardata dall'esistenza di buchi neri formati da galassie in fase di collasso. In teoria l'energia potrebbe essere estratta dal campo gravitazionale di un buco nero; è quindi ammissibile che una civiltà progredita possa mantenersi per almeno $10^{100}$ anni, cioè per il tempo necessario ai buchi neri per evaporare. Lo schema conosciuto più efficace per estrarre energia da un buco nero è quello proposto da Charles W. Misner dell'Università del Maryland di College Park, da Kip S. Thorne del California Institute of Technology e da John Archibald Wheeler dell'Università del Texas ad Austin. Attorno a un buco nero in rotazione viene costruito un guscio rigido e il materiale di scarto della civiltà viene spedito con un satellite vicino all'orizzonte degli eventi, il confine della regione dalla quale nulla può sfuggire. Nei pressi dell'orizzonte degli eventi il materiale viene scaricato all'equatore del buco nero. Il risultato è che tanto l'energia potenziale gravitazionale del materiale scaricato quanto l'energia rotazionale del buco nero sono diminuite. Dato che l'energia totale del satellite e del buco nero deve conservarsi, l'energia del satellite aumenta di una quantità pari alla somma delle due diminuzioni energetiche. In teoria l'energia acquistata dal satellite potrebbe trasformarsi in una forma utile.**




**L'universo può essere ciclico, se la densità della materia è sufficiente a chiuderlo e i processi fisici finora sconosciuti permettono all'universo di «invertire la rotta», cioè di espandersi ancora. Dato che durante la fase di contrazione si acquista energia, ogni nuovo ciclo ha una fase di espansione più lunga della precedente. Il periodo di espansione dovrebbe aumentare di un fattore due circa durante il prossimo ciclo dell'universo, supponendo che l'espansione attuale continui per circa 100 volte l'attuale età dell'universo. La durata dei cicli successivi aumenta di fattori maggiori.**

nia Institute of Technology, hanno cominciato a speculare su temi come il fabbisogno energetico per una conservazione indefinita della vita e per le comunicazioni tra i vari membri di una società dispersa in regioni di spazio con dimensioni sempre più grandi.

Secondo Dyson la vita e la coscienza non sono necessariamente presenti solo all'interno delle cellule e nel DNA; anzi, egli sostiene, l'aspetto essenziale della coscienza è la complessità di struttura, che può essere realizzata in qualsiasi materiale di natura opportuna. Così egli sostiene che l'idea di un calcolatore che sente o di una nuvola che sente non si può semplicemente scartare come incoerente da un punto di vista filosofico.

Con tali ipotesi, le variazioni nell'ambiente causate dalla morte e dal raffreddamento delle stelle e dalla loro evaporazione dalle galassie non dovrebbero essere insormontabili per un sistema che può essere considerato vivente e intelligente. Per esempio, in linea di principio, si potrebbe estrarre energia dal campo gravitazionale di un buco nero supermassivo. Tuttavia il decadimento di protoni e neutroni può portare a un cambiamento fondamentale, dato che sembra improbabile (anche se forse non del tutto impossibile) che l'intelligenza possa basarsi sulle deboli fondamenta di elettroni e positoni. Inoltre, se l'universo è chiuso, le condizioni necessarie perché possa realizzarsi la vita possono esistere soltanto per certi periodi durante ogni ciclo.

In un universo aperto l'estremo impedimento alla vita è del tutto diverso. Con l'evaporazione dei buchi neri si verificherà una carenza di energia cosmica, perché, mentre continua l'espansione dell'universo, le particelle e i fotoni restanti perdono energia. Qualsiasi tasso costante di consumo di energia da parte delle forme di vita diventerà a tempo debito insostenibile. D'altra parte, Dyson avanza l'idea che, ai periodi di consumo, potrebbero alternarsi periodi sempre più lunghi di ibernazione, durante i quali non viene consumata energia. In questo universo esiste quindi la possibilità di civiltà di durata straordinariamente lunga.

Il punto di partenza della cosmologia contemporanea, come abbiamo sottolineato, è la teoria del big bang. La nostra trattazione quantitativa dell'evoluzione del big bang nel futuro remoto si basa sulle nuove teorie che descrivono la fisica delle particelle elementari. Queste idee però erano già state espresse nel loro spirito dall'astronomo e poeta persiano del secolo XI Omar Khayyām, il cui poema *Rubâiyât* è stato tradotto dallo scrittore inglese Edward FitzGerald. In una quartina Omar scrisse:

Con il primo fango della terra fecero
l'impasto dell'ultimo uomo,
E poi gettarono il seme
dell'ultima messe:
Sì, il primo giorno della creazione
scrisse
Ciò che l'ultima alba del giorno del giudizio
leggerà.

# Reazioni chimiche oscillanti

*Un tempo ritenute contrarie alle leggi naturali, è stato invece possibile realizzarle di recente in laboratorio; sembra che possano chiarire analoghi comportamenti nei sistemi viventi*

di Irving R. Epstein, Kenneth Kustin, Patrick De Kepper e Miklós Orbán

Ovunque in fisica, in astronomia e in biologia si trovano fenomeni oscillanti, o periodici. Essi vanno dal familiare movimento del pendolo e delle orbite dei pianeti ai complessi orologi biologici che governano il comportamento giornaliero e stagionale degli organismi viventi. Fino ad epoca recente, invece, i chimici credevano che le reazioni che avvengono nelle provette e nei bicchieri da laboratorio fossero singolarmente immuni da quell'andamento periodico che è così familiare in altri settori della scienza. Infatti la maggior parte dei chimici formatisi prima del 1950 era probabilmente convinta che, per una miscela di sostanze inorganiche semplici, il partecipare a una reazione che oscillava in modo visibile e periodico sarebbe stato una evidente violazione di una immutabile legge di natura. Anche oggi si suppone abitualmente che le reazioni chimiche siano unidirezionali: se due sostanze reagiscono per darne una terza, ci si aspetta che la reazione continuerà regolarmente fino all'esaurimento dei reagenti o al raggiungimento di uno stato di equilibrio. Normalmente non ci si aspetta che le concentrazioni dei prodotti intermedi della reazione raggiungano un certo livello, quindi scendano a un livello inferiore per risalire poi e di nuovo abbassarsi ripetutamente, finché si formano prodotti stabili resistenti a successivi cambiamenti.

Benché resoconti di tali reazioni apparissero sporadicamente nella letteratura chimica della fine dell'Ottocento e degli inizi del Novecento, essi furono respinti dalla larga maggioranza dei chimici come fenomeni non riproducibili, che potevano essere attribuiti probabilmente a processi estranei, come la corrosione, o la formazione di pellicole, che avvenivano nel corso della reazione. Negli ultimi 25 anni, tuttavia, lo studio delle reazioni chimiche oscillanti ha finalmente conquistato rispettabilità e ora costituisce uno dei campi della chimica in più rapido sviluppo. Lo studio promette di penetrare nella dinamica chimica e nei meccanismi di catalisi e forse nei fenomeni periodici ancora oscuri che si osservano in biologia e in geologia. Mentre il primitivo piccolo gruppo di reazioni chimiche oscillanti fu trovato per caso, è ora possibile descrivere le condizioni che contribuiscono all'oscillazione e preparare una ricerca sistematica di nuovi sistemi oscillanti. Si tratta di una ricerca che ha già dato i suoi frutti.

La riluttanza dei chimici ad accettare la realtà delle reazioni oscillanti può essere fatta risalire principalmente al secondo principio della termodinamica. Nella sua formulazione più nota, dovuta al fisico tedesco Rudolf Clausius nel XIX secolo, il secondo principio stabilisce che l'entropia, ossia la casualità, dell'universo tende ad aumentare. Applicato alle reazioni chimiche, il principio richiede che un sistema chimico, in assenza di apporti esterni di materia o di energia, deve continuamente tendere a uno stato finale di equilibrio. Cioè, se *A* va verso *B*, deve essere così per tutta la durata della reazione senza deviazioni a ritroso verso *A* lungo il percorso. Si pensava che le reazioni in cui questa regola sembra essere violata dovessero derivare o da uno scarso controllo delle condizioni sperimentali, o da qualche artificio voluto, dal momento che esse costituirebbero una sorta di macchina chimica a moto perpetuo.

Si può pertanto immaginare l'indifferenza con cui fu accolta la relazione su una reazione oscillante pubblicata nel 1921 da William C. Bray dell'Università della California a Berkeley. Studiando il ruolo dello iodato, ione dello iodio ossigenato ($IO_3^-$), come catalizzatore della decomposizione dell'acqua ossigenata, o perossido di idrogeno, in acqua e ossigeno, Bray osservò che, in determinate condizioni, la velocità di produzione dell'ossigeno e la concentrazione dello iodio nella soluzione cambiavano periodicamente. All'osservazione di Bray fu prestata scarsa attenzione nel corso dei successivi cinquant'anni. I pochi articoli che apparvero sulla reazione di Bray tentavano in gran parte di spiegare le oscillazioni come artefatti dovuti a polvere o a impurità più che di esaminare ulteriormente il fenomeno.

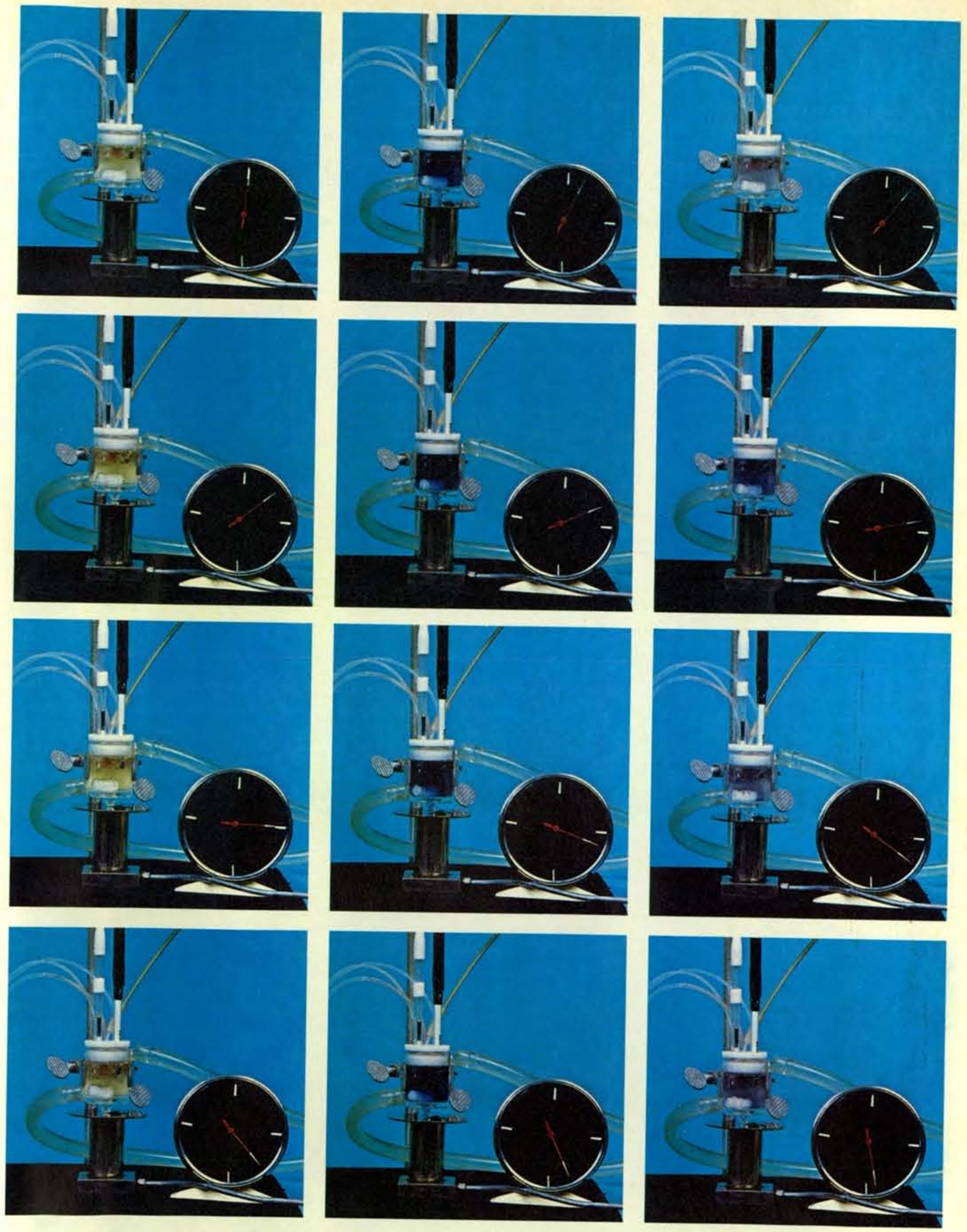
Come vedremo, lo iodio e gli ioni ioduro ($I^-$) e iodato svolgono un ruolo importante in molti sistemi oscillanti messi a punto di recente. L'era moderna per quanto riguarda lo studio di reazioni chimiche oscillanti può essere fatta partire da una scoperta accidentale dovuta nel 1958 al chimico sovietico B. P. Belousov. Egli osservò che, se si sciolgono in acqua acido citrico e acido solforico con bromato di potassio e un sale di cerio, la soluzione varia periodicamente da incolore a color giallo chiaro. Col 1958, dunque, alcuni chimico-fisici, almeno, furono disposti a considerare seriamente la notizia, sia perché la reazione di Belousov era facilmente ripetibile, sia perché i concetti di termodinamica del XIX secolo erano stati estesi dopo la seconda guerra mondiale.

Il primo in questo impegno fu Ilya Prigogine della Université libre di Bruxelles, il quale riconobbe che la termodinamica classica di Clausius richiedeva non solo che i sistemi fossero isolati dall'ambiente, ma anche che fossero vicini al loro stato di equilibrio. Per sistemi lontani dall'equilibrio, sia perché una reazione è solo ai primi stadi, sia perché il sistema è «aperto», e pertanto soggetto a un flusso di energia dall'esterno, Prigogine e i suoi collaboratori svilupparono il concetto di una termodinamica irreversibile. Per questo lavoro, nel 1977, Prigogine ha ricevuto il premio Nobel per la chimica.

In sistemi lontani dall'equilibrio, possono presentarsi molti nuovi fenomeni, descritti come strutture dissipative. Tali strutture comprendono oscillazioni periodiche nelle concentrazioni di specie intermedie in una reazione chimica; invece, non sono soggetti a oscillazione né i reagenti iniziali, né i prodotti finali. I viventi costituiscono gli esempi più interessanti e vari di sistemi oscillanti aperti, lontani dall'equilibrio. Essi vengono mantenuti in uno stato di non equilibrio dall'apporto di reagenti (sostanze nutritive) provenienti dall'esterno e dalla produzione di sostanze di rifiuto. Se l'uno o l'altro di questi flussi si arresta, l'organismo e le sue oscillazioni si estinguono.

Nel 1958, le implicazioni della termodinamica irreversibile di Prigogine furono





**Un sistema chimico oscillante è fotografato nel laboratorio degli autori alla Brandeis University. La reazione procede in un recipiente di vetro nel quale vengono pompate tre soluzioni a una velocità costante: iodato di potassio, acido perclorico e acqua ossigenata, e acido malonico con solfato manganoso monoidrato. L'amido serve da indicatore, poiché in presenza di iodio forma un complesso blu scuro. L'intervallo fra una fotografia e la successiva, partendo dall'alto a sinistra e procedendo da sinistra verso destra, è poco meno di tre secondi (come è indicato dalla lancetta dei secondi sull'orologio). Questo particolare sistema è stato progettato da due insegnanti di scuola superiore in California, Thomas S. Briggs e Warren C. Rauscher. La reazione è stata modificata per farla avvenire in condizioni di flusso continuo.**

capite soltanto da pochi chimici. Inoltre, la scoperta di Belousov passò quasi inosservata poiché era stata pubblicata in una raccolta sovietica poco nota di estratti tecnici sul ruolo delle radiazioni in medicina. La reazione oscillante di Belousov presentava importanti vantaggi rispetto al primitivo sistema di Bray. Essa avveniva a temperatura ambiente e non generava prodotti nocivi. Inoltre, le oscillazioni erano chiaramente visibili quando il cerio passava continuamente da uno stato ossidato (cerico), di colore giallo, a uno stato meno ossidato (ceroso), che è incolore.

Pochi anni dopo la sua pubblicazione, la reazione di Belousov attirò l'attenzione di A. M. Zhabotinsky dell'Istituto di fisica biologica nei pressi di Mosca. Egli apportò modifiche ai reagenti di Belousov (sostituendo, ad esempio, il sale di cerio con un sale di ferro che mostra un viraggio di colore dal rosso al blu, più evidente) e iniziò uno studio sistematico della reazione, che ora è comunemente chiamata reazione BZ (da Belousov-Zhabotinsky). All'inizio degli anni sessanta Zhabotinsky pubblicò un'ampia messe di risultati sperimentali, comprendenti la scoperta che, se si lascia indisturbato uno strato sottile di soluzione rossa BZ, inizialmente uniforme, compaiono ben presto dei punti blu che si aggregano fino ad assumere una sorprendente disposizione di strutture ad anello.

La conoscenza della reazione BZ (e della termodinamica irreversibile di Prigogine) si diffuse rapidamente sul finire degli anni sessanta e all'inizio degli anni settanta. Benché i chimici di tutto il mondo fossero affascinati dal fenomeno BZ, la grande maggioranza lo considerava ancora una curiosità, più utile per sorprendere i partecipanti a convegni, che come oggetto di studi seri. In breve tempo furono sviluppate molte varianti della reazione BZ, sostituendo uno o più componenti con specie chimiche molto affini per ottenere una reazione oscillante più lenta, più veloce, o diversamente colorata. Tra le varianti, quella che visivamente colpì di più fu un ibrido dei sistemi BZ e Bray, messo a punto nel 1973 da due insegnanti di una scuola superiore di San Francisco, Thomas S. Briggs e Warren C. Rauscher. Il loro sistema, che comprende acqua ossigenata, iodato di potassio, acido perclorico, acido malonico, solfato di manganese e amido, varia da incolore, al colore oro e al blu, per poi ricominciare da capo.

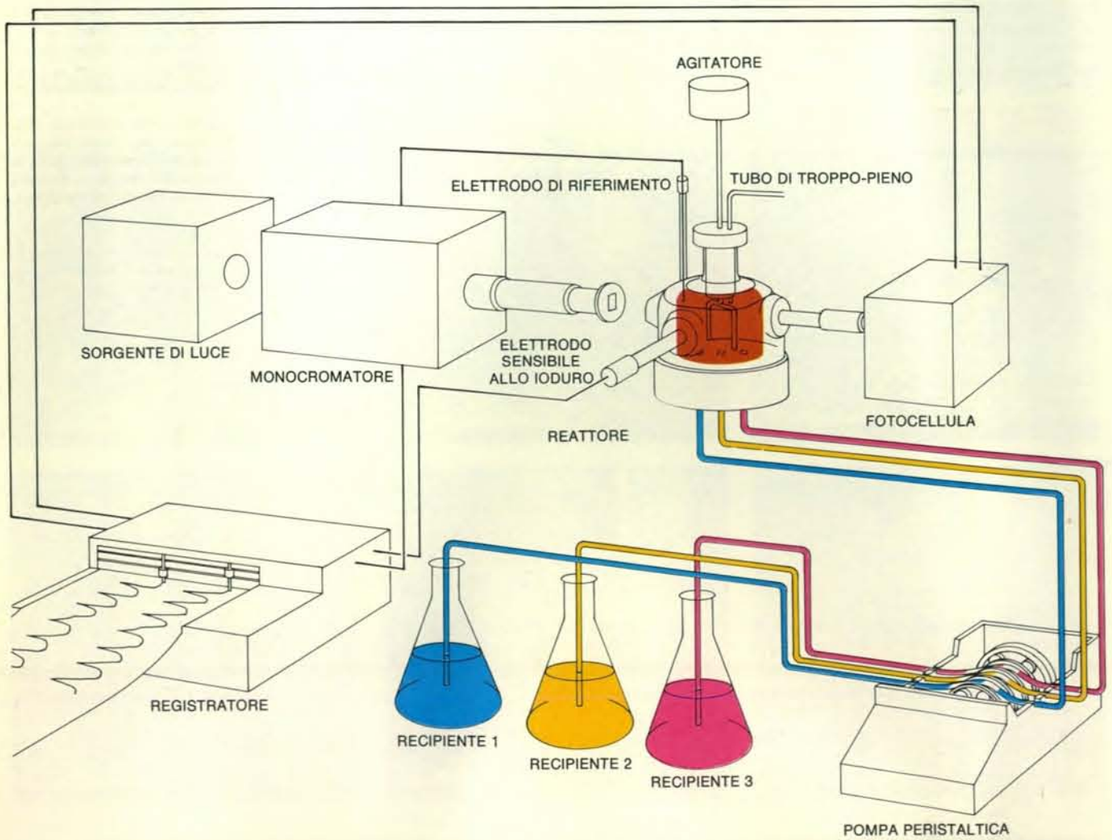
Benché la reazione BZ e le altre da questa derivate fossero diventate ben note verso la fine degli anni sessanta, grazie a dimostrazioni effettuate in ambiente scientifico, la natura del fenomeno oscillatorio restava un mistero. Un chimico capisce una reazione solo quando può scriverne il meccanismo, cioè una serie di reazioni, chiamate stadi elementari, ciascuna delle quali descrive un reale incontro tra molecole. La familiare equazione «stechiometrica», che i chimici sono soliti scrivere per descrivere una reazione, mostra solo il risultato netto e non il meccanismo con cui la reazione procede effettivamente su scala molecolare. Ad esempio, quando l'acetilene ($C_2H_2$) si combina con l'ossigeno ($O_2$) in un cannello per saldatura per formare anidride carbonica ($CO_2$) e acqua ($H_2O$), la reazione globale può essere scritta come $2\ C_2H_2 + 5\ O_2 \rightarrow 4\ CO_2 + 2\ H_2O$.

È però molto improbabile che due molecole di acetilene e cinque molecole di ossigeno vengano a collidere simultaneamente e si scindano, dando origine a quattro molecole di anidride carbonica e a due di acqua. La reazione in realtà avviene mediante una sequenza di stadi elementari, ciascuno dei quali comporta una collisione tra due molecole, o la scissione di una molecola singola, spesso con formazione o distruzione di specie intermedie che non compaiono affatto nell'equazione della reazione netta. Chiarire un tale meccanismo rappresenta una enorme sfida che richiede ingegnosità, l'analisi di molti dati sperimentali e spesso un'estesa simulazione al calcolatore.

Agli inizi degli anni settanta, si assunsero il formidabile compito di costruire un meccanismo per la reazione oscillante BZ Richard M. Noyes dell'Università dell'Oregon, in collaborazione con Richard J. Field (che è ora all'Università del Montana), ed Endre Körös dell'Università Eötvös di Budapest. Entro il 1972 essi riuscirono a ideare uno schema composto di 18 stadi elementari, che comportava circa 20 specie chimiche diverse e che sembrava in grado di spiegare le oscillazioni. Due anni dopo, una simulazione dettagliata al calcolatore svolta da Noyes, Field e David Edelson dei Bell Laboratories confermò che quel meccanismo prevede le oscillazioni. Studi successivi hanno dimostrato che esso spiega bene anche lo sviluppo di strutture spaziali. Noyes ha paragonato il suo metodo per chiarire il meccanismo al «metodo di [Sherlock] Holmes: "Quando le altre possibilità vengono meno, ciò che rimane, *per quanto improbabile,* deve essere la verità"».

Con la soluzione del meccanismo BZ, le reazioni chimiche oscillanti divennero un'area importante e stimolante per coloro che erano interessati ai meccanismi di reazione e alla dinamica chimica. Furono avanzate varie ipotesi sulle condizioni richieste perché ci sia oscillazione in una reazione chimica e in alcuni ambienti nacque l'ottimistica speranza che la comprensione dell'oscillazione chimica potesse gettare nuova luce sul comportamento periodico degli organismi viventi. Nessuno, tuttavia, riuscì a definire una serie di condizioni necessarie e sufficienti per l'oscillazione chimica. Quando gli anni settanta giunsero al termine, il numero di reazioni chimiche diverse con andamento oscillante e allora note era terribilmente scarso; tutto era stato scoperto per caso; si conosceva una sola reazione e nessuno sapeva come scoprirne altre.

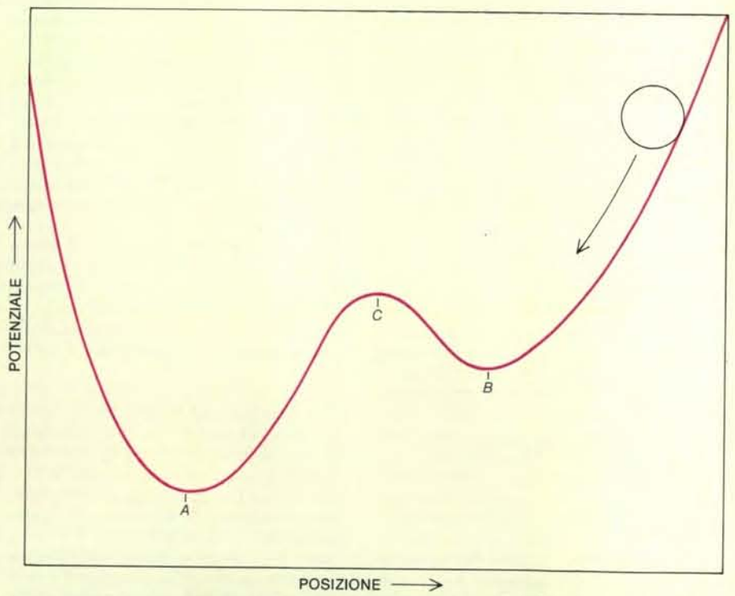
Anche se i sistemi biologici sono notoriamente una fonte fertile di preparati che mostrano oscillazione, essi tendono a sottrarsi al tipo di analisi meccanicistica che porta a una generalizzazione. Tali sistemi oscillanti possono essere ottenuti, per esempio, estraendo il contenuto chimico delle cellule di lievito e separandolo dalle pareti cellulari e da altre strutture estra-



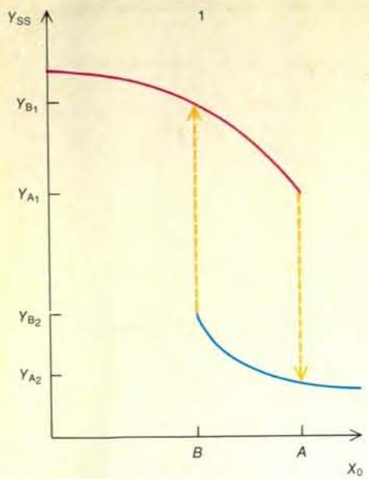

**Il reattore a flusso e agitazione continui ha facilitato la ricerca degli autori sulle reazioni chimiche oscillanti, fornendo le condizioni di non equilibrio necessarie perché abbiano luogo tali reazioni. L'illustrazione mostra una tipica disposizione sperimentale. I reagenti, contenuti in tre recipienti, vengono pompati in modo continuo alla base del recipiente di reazione. Nel primo sistema oscillante scoperto dagli autori, i materiali introdotti erano soluzioni acquose acide di un clorito, di uno iodato e di un arsenito. Le variazioni periodiche di colore sono controllate mediante la fotocellula. Le oscillazioni nella concentrazione di ioni particolari (in questo caso ioduri) sono rilevate da un elettrodo.**

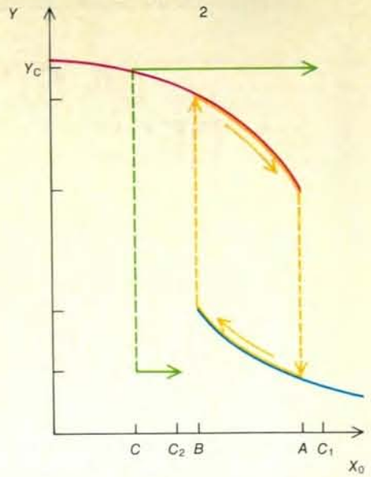
| | |
|---|---|
| 1 | $2H^+ + Br^- + BrO_3^- \rightleftharpoons HOBr + HBrO_2$ |
| 2 | $H^+ + HBrO_2 + Br^- \rightleftharpoons 2HOBr$ |
| 3 | $HOBr + Br^- + H^+ \rightleftharpoons Br_2 + H_2O$ |
| 4 | $CH_2(COOH)_2 \rightleftharpoons (OH)_2C{=}CHCOOH$ |
| 5 | $Br_2 + (OH)_2C{=}CHCOOH \rightleftharpoons H^+ + Br^- + BrCH(COOH)_2$ |
| 6 | $HBrO_2 + BrO_3^- + H^+ \rightleftharpoons 2BrO_2 + H_2O$ |
| 7 | $BrO_2 + Ce^{3+} + H^+ \rightleftharpoons Ce^{4+} + HBrO_2$ |
| 8 | $Ce^{4+} + BrO_2 + H_2O \rightleftharpoons BrO_3^- + 2H^+ + Ce^{3+}$ |
| 9 | $2HBrO_2 \rightleftharpoons HOBr + BrO_3^- + H^+$ |
| 10 | $Ce^{4+} + CH_2(COOH)_2 \rightleftharpoons CH(COOH)_2 + Ce^{3+} + H^+$ |
| 11 | $CH(COOH)_2 + BrCH(COOH)_2 + H_2O \rightleftharpoons Br^- + CH_2(COOH)_2 + HOC(COOH)_2 + H^+$ |
| 12 | $Ce^{4+} + BrCH(COOH)_2 + H_2O \rightleftharpoons Br^- + HOC(COOH)_2 + Ce^{3+} + 2H^+$ |
| 13 | $2HOC(COOH)_2 \rightleftharpoons HOCH(COOH)_2 + O{=}CHCOOH + CO_2$ |
| 14 | $Ce^{4+} + HOCH(COOH)_2 \rightleftharpoons HOC(COOH)_2 + Ce^{3+} + H^+$ |
| 15 | $Ce^{4+} + O{=}CHCOOH \rightleftharpoons O{=}CCOOH + Ce^{3+} + H^+$ |
| 16 | $2O{=}CCOOH + H_2O \rightleftharpoons O{=}CHCOOH + HCOOH + CO_2$ |
| 17 | $Br_2 + HCOOH \rightarrow 2Br^- + CO_2 + 2H^+$ |
| 18 | $2CH(COOH)_2 + H_2O \rightleftharpoons CH_2(COOH)_2 + HOCH(COOH)_2$ |

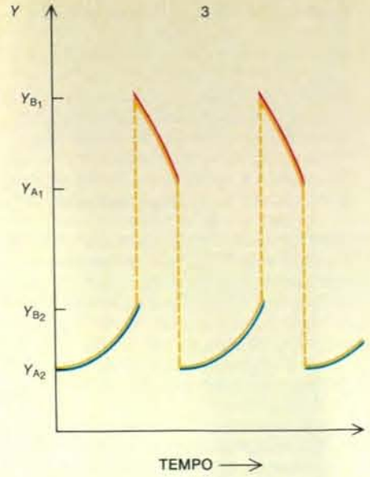
**La simulazione al calcolatore dell'oscillazione nella reazione di Belousov-Zhabotinsky comporta 18 stadi elementari e 21 specie chimiche diverse. La reazione prende il nome da due chimici sovietici: B. P. Belousov, al quale si deve la sua scoperta, e A. M. Zhabotinsky, che la perfezionò. I materiali di partenza sono tre sostanze inorganiche, gli ioni bromato ($BrO_3^-$), bromuro ($Br^-$) e cerosi ($Ce^{3+}$) oltre a una sostanza organica, l'acido malonico ($CH_2(COOH)_2$). Un mezzo a base di acido solforico fornisce gli ioni idrogeno ($H^+$). I prodotti della reazione sono anidride carbonica ($CO_2$), acido formico (HCOOH) e acido bromomalonico ($BrCH(COOH)_2$). Dato che il cerio oscilla fra i due stati di ossidazione cerico ($Ce^{4+}$) e ceroso ($Ce^{3+}$), la soluzione vira alternativamente da gialla a incolore. Le reazioni rappresentate in colore comportano solo specie inorganiche e sono meglio conosciute delle reazioni in nero, le quali coinvolgono specie derivate dall'acido malonico. Questo meccanismo è stato messo a punto all'Università dell'Oregon da Richard M. Noyes con il collega Richard J. Field e Endre Körös dell'Università Eötvös di Budapest.**




**Il sistema chimico bistabile può essere rappresentato da un'analogia meccanica. Una palla che rotola in una buca di potenziale con due punti minimi, *A* e *B*, può fermarsi in uno dei due. *A* e *B* rappresentano pertanto gli stati stazionari stabili del sistema; *C* è uno stato stazionario instabile. Una palla può fermarsi in *C*, ma la più piccola perturbazione la farà scivolare in *A* o in *B*. Qui la variabile è la posizione della palla; il vincolo esterno è rappresentato dalla forma della buca di potenziale.**




**Il ciclo di isteresi è illustrato in un sistema chimico bistabile, che opera in un reattore a flusso e agitazione continui. Quando la concentrazione $X_0$ della sostanza $X$ che fluisce nel reattore aumenta, il valore dello stato stazionario $Y_{SS}$ della specie $Y$ nel reattore decresce lentamente lungo la curva in rosso finché, al punto critico $A$, cade da $Y_{A1}$ a $Y_{A2}$. Se $X_0$ viene ulteriormente aumentata, $Y_{SS}$ cade ancora, ma lungo la curva in blu. Se $X_0$ diminuisce, $Y_{SS}$ ripercorre la curva in blu al di là del punto $A$ finché raggiunge un secondo punto critico $B$. Qui $Y_{SS}$ balza da $Y_{B2}$ a $Y_{B1}$. Tra $A$ e $B$ può esistere l'uno o l'altro stato stazionario.**

**Può verificarsi oscillazione se una adatta sostanza $Z$ viene aggiunta a un sistema chimico bistabile che presenta isteresi. Se $Z$ reagisce con la sostanza $Y$ a bassa velocità per dare la sostanza $X$, la concentrazione effettiva di quest'ultima *($X_0$)* viene modificata in modo che il sistema «veda» un valore di $X$ più elevato di quello fornito dal recipiente *(a sinistra)*. L'effetto di una data quantità di $Z$ su $X_0$ *(frecce in verde)* è maggiore nel tratto rosso del ciclo di isteresi (dove c'è più $Y$) rispetto al tratto blu. Immaginiamo che la reazione inizi con un flusso di $X$ uguale a $C$. Ora aggiungiamo abbastanza $Z$, una quantità $Z_0$, così che l'effettiva $X_0$ sia uguale a $C_1$ quando il sistema è nel tratto rosso e a $C_2$ quando è nel tratto blu. In assenza di $Z$, il flusso $X_0 = C$ darebbe un valore di stato stazionario $Y_C$. Invece, con l'aggiunta di $Z_0$ il sistema segue il tratto rosso quando il valore effettivo di $X$ aumenta lentamente per effetto della reazione di $Y$ con $Z$. Il sistema tenta di raggiungere il valore di $C_1$, ma nel raggiungere $A$ cade dal tratto rosso e subisce una rapida transizione al tratto blu. A questo punto, il sistema «vede» un valore di $X_0 = C_2$ e così devìa verso sinistra lungo il tratto blu nel tentativo di adattare il nuovo effettivo valore di $X_0$. Prima che raggiunga lo scopo, a $X_0 = B$ salta di nuovo al tratto rosso e il ciclo è completo *(frecce in giallo)*:**

nee. Il «brodo» ottenuto, quando viene alimentato con appropriate sostanze nutritive, mostra la sua oscillazione nella concentrazione degli ioni idrogeno e della molecola di nicotinammideadenindinucleotide, che nella sua forma ridotta (NADH) trasporta elettroni nel normale ciclo energetico della cellula. Alla Johnson Research Foundation, Britton Chance e collaboratori hanno dimostrato che molti sistemi oscillanti biologici di questo tipo comportano la glicolisi, il processo mediante il quale le cellule producono energia metabolizzando zucchero in assenza di ossigeno. Poiché tali reazioni sono catalizzate da molti enzimi, la loro complessità è davvero scoraggiante.

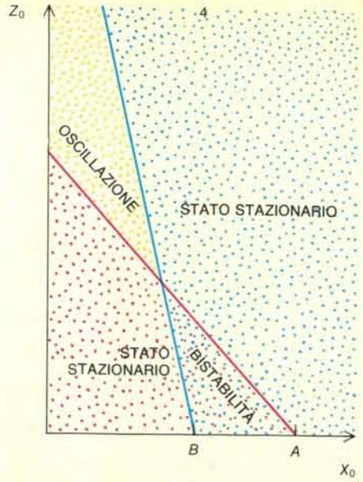
Nel 1979 due di noi (Epstein e Kustin) hanno deciso che era giunto il momento di tentare di ideare una procedura sistematica per mettere a punto un sistema chimico oscillante. Ma da dove si doveva incominciare? Siamo partiti identificando tre condizioni che sapevamo essere necessarie per l'oscillazione chimica, o che contribuivano ad essa. La prima condizione è che i sistemi chimici possono oscillare solo se sono lontani dall'equilibrio. La seconda è la retroazione *(feedback)*: parte del prodotto di uno stadio della sequenza di reazioni deve esercitare un'influenza sulla sua stessa velocità di formazione. La terza è che il sistema chimico deve mostrare bistabilità. Nelle stesse condizioni esterne, definite vincolanti, il sistema deve poter esistere in due differenti stati stazionari stabili.

Il requisito che il sistema sia mantenuto lontano dall'equilibrio potrebbe essere soddisfatto facendo avvenire la reazione in un apparecchio ben noto agli ingegneri chimici: il reattore a flusso e agitazione continui (CSTR, *continuous-flow stirred-tank reactor)*. Alla metà degli anni settanta, Adolphe Pacault e collaboratori del Centre de Recherches Paul Pascal di Bordeaux hanno adattato il CSTR allo studio di reazioni chimiche oscillanti. Quando, nel 1980, uno di noi (De Kepper) ha lasciato Bordeaux per raggiungere il nostro gruppo alla Brandeis University, è parso naturale costruire un reattore di questo tipo. L'arrivo dall'Ungheria di Orbán, che aveva lavorato a lungo su sistemi oscillanti a base di bromato, ha completato il nostro gruppo.

Un tipo comune di retroazione incontrata nei sistemi viventi è l'autocatalisi, in cui la velocità con la quale una sostanza è prodotta aumenta con la concentrazione. Il concetto si estende a popolazioni di organismi: il numero di organismi aggiunti a una popolazione in un dato intervallo è quasi sempre proporzionale al numero di individui già presenti nella popolazione. (Questa osservazione mette in rilievo l'affermazione, fatta da Thomas Malthus nel 1798, secondo la quale le popolazioni tendono a esaurire le scorte di cibo.) In chimica i sistemi autocatalitici sono rari, ma non straordinari. Già nel 1920 Alfred J. Lotka della Johns Hopkins University dimostrò che un semplice schema, costituito da due reazioni autocatalitiche accoppiate, avrebbe dato luogo a oscillazione. Il meccanismo di Lotka, pur essendosi dimostrato che non era applicabile ad alcuna reazione chimica effettiva, è servito da guida per molti chimici ed è stato di considerevole utilità nel descrivere le oscillazioni di popolazioni di predatori e di prede nei sistemi ecologici.

Consideriamo ora il ruolo della bistabilità, ossia l'esistenza di due differenti stati stazionari stabili in un sistema chimico oscillante. Per stato stazionario si intende una condizione del sistema in cui tutte le variabili, come la concentrazione di ciascuna specie chimica, hanno raggiunto valori costanti. Lo stato stazionario è stabile se può adattarsi a piccoli cambiamenti di una variabile, come l'aggiunta di una goccia di acido, rimanendo tale senza essere trasformato in un nuovo stato. Se la più piccola variazione provoca una transizione a un differente stato del sistema, si dice che lo stato originario è instabile. Per esempio, una piccola palla sul fondo di un recipiente si trova in uno stato stabile. Una palla in equilibrio sul bordo del recipiente è in uno stato instabile. In un reattore a flusso e agitazione continui le condizioni vincolanti esterne che controllano i valori dello stato stazionario sono generalmente la temperatura del bagno ter-






**ora $X_0$ ha nuovamente il valore effettivo $C_1$, così che il sistema si muove ancora verso destra. La ripetizione di questa sequenza dà origine a un'oscillazione periodica nel valore di $Y$ *(al centro)*. Il diagramma a più fasi *(a destra)* mostra come l'innalzamento di $Z_0$ può spostare il sistema dalla bistabilità all'oscillazione. Se $Z$ è uguale a 0, il sistema mostra due condizioni di stato stazionario: lo stato «rosso» per un basso valore di $X_0$ e lo stato «blu» per un valore elevato di $X_0$. Quando $X_0$ viene a cadere tra $A$ e $B$, il sistema è bistabile. Con l'aumentare di $Z_0$, la regione di bistabilità si stringe finché a un valore critico inizia l'oscillazione.**

mostatico, le concentrazioni dei reagenti chimici nel serbatoio di alimentazione del reattore e le loro velocità di flusso.

Se vengono cambiate le condizioni vincolanti esterne in un sistema chimico bistabile, può verificarsi un fenomeno più particolare, noto come isteresi. Questo fenomeno è comune nel magnetismo, dove assume la forma di un «ciclo di isteresi». Quando un pezzo di ferro è sottoposto a un aumento di forza magnetica, esso raggiunge alla fine la completa saturazione magnetica. Se la forza magnetica si riduce poi a zero, il ferro conserva parte della sua magnetizzazione. Al fine di ridurre anche questa a zero, il campione di ferro va sottoposto a una forza magnetica in senso opposto. Il ciclo completo di magnetizzazione e smagnetizzazione prende l'aspetto di una *S*, spessa al centro e appuntita alle estremità.

Un fenomeno analogo può verificarsi in un sistema chimico bistabile. Nel 1979 un calcolo effettuato da Jacques Boissonade e da uno di noi (De Kepper) al Centre de Recherches Paul Pascal suggerì che le oscillazioni potrebbero essere indotte aggiungendo un altro composto chimico a un sistema capace di isteresi. Il comportamento di un sistema di questo tipo può essere rappresentato in un diagramma a più fasi, il quale mostra che, se la sostanza aggiunta esercita effetti diversi sui due rami degli stati stabili del sistema, è probabile che si sviluppino oscillazioni.


# LA SCIENZA DEI CALCOLATORI

*Sul continuo progredire della microelettronica e sulle prestazioni sempre più sofisticate e complesse dei calcolatori*

**LE SCIENZE**

*edizione italiana di* **SCIENTIFIC AMERICAN** *ha pubblicato numerosi articoli tra cui:*

**MONTAGGIO GUIDATO DAL CALCOLATORE**
di J. L. Nevins e D. E. Whitney (n. 116)

Nella produzione in grande serie il montaggio viene eseguito a mano o con macchine specializzate, mentre automatismi programmabili possono essere economicamente convenienti per il montaggio in piccola serie.

**LINGUAGGI DI PROGRAMMAZIONE**
di A. Feldman (n. 138)

I linguaggi di programmazione ad alto livello hanno trasformato la elaborazione delle informazioni, fornendo meccanismi per codificare problemi complessi in modo da renderli risolubili con il calcolatore.

**IL CALCOLATORE A SUPERCONDUZIONE**
di J. Matisoo (n. 143)

Una nuova tecnologia microelettronica consentirebbe di sostituire superconduttori ai transistori. Con tali componenti i calcolatori potrebbero eseguire un miliardo di operazioni elementari al secondo.

**BACKGAMMON AL CALCOLATORE**
di H. Berliner (n. 144)

Il backgammon costituisce un buon banco di prova per lo sviluppo della intelligenza artificiale. *BKG 9.8* è il primo programma di calcolo che abbia battuto un campione mondiale in un gioco di carte o da tavolo.

**LA TECNOLOGIA DELLE MEMORIE A DISCO**
di R. M. White (n. 146)

Sulla superficie di un disco rotante si possono immagazzinare dati per un calcolatore, in forma di piccole areole magnetizzate. In dispositivi in corso di sviluppo i dati verranno «scritti» e «letti» da un laser.

**ELABORAZIONE DI IMMAGINI CON IL CALCOLATORE**
di T. M. Cannon e B. R. Hunt (n. 160)

Quando le informazioni contenute in un'immagine sono espresse in forma digitale, possono essere elaborate meglio matematicamente che non per via ottica. Con questo sistema è possibile rendere nitida una fotografia sfocata.

**ALGEBRA AL CALCOLATORE**
di R. Pavelle, M. Rothstein e J. Fitch (n. 162)

Un calcolatore può manipolare non soltanto numeri, ma anche simboli. Nuovi algoritmi di uso generale possono svolgere molti compiti matematici ripetitivi oppure risolvere problemi altrimenti intrattabili.

**SUPERCALCOLATORI**
di R. D. Levine (n. 163)

Il Cray-1 e il CYBER 205 possono eseguire 100 milioni di operazioni aritmetiche al secondo: questa incredibile velocità è necessaria per risolvere problemi complessi come per esempio quelli della dinamica dei fluidi.

**INTELLIGENZA ARTIFICIALE**
di D. L. Waltz (n. 172)

Alcuni programmi di calcolatore possono giocare, elaborare informazioni visive, imparare dall'esperienza e capire frammenti di linguaggio, ma non sono in grado di simulare il buon senso.

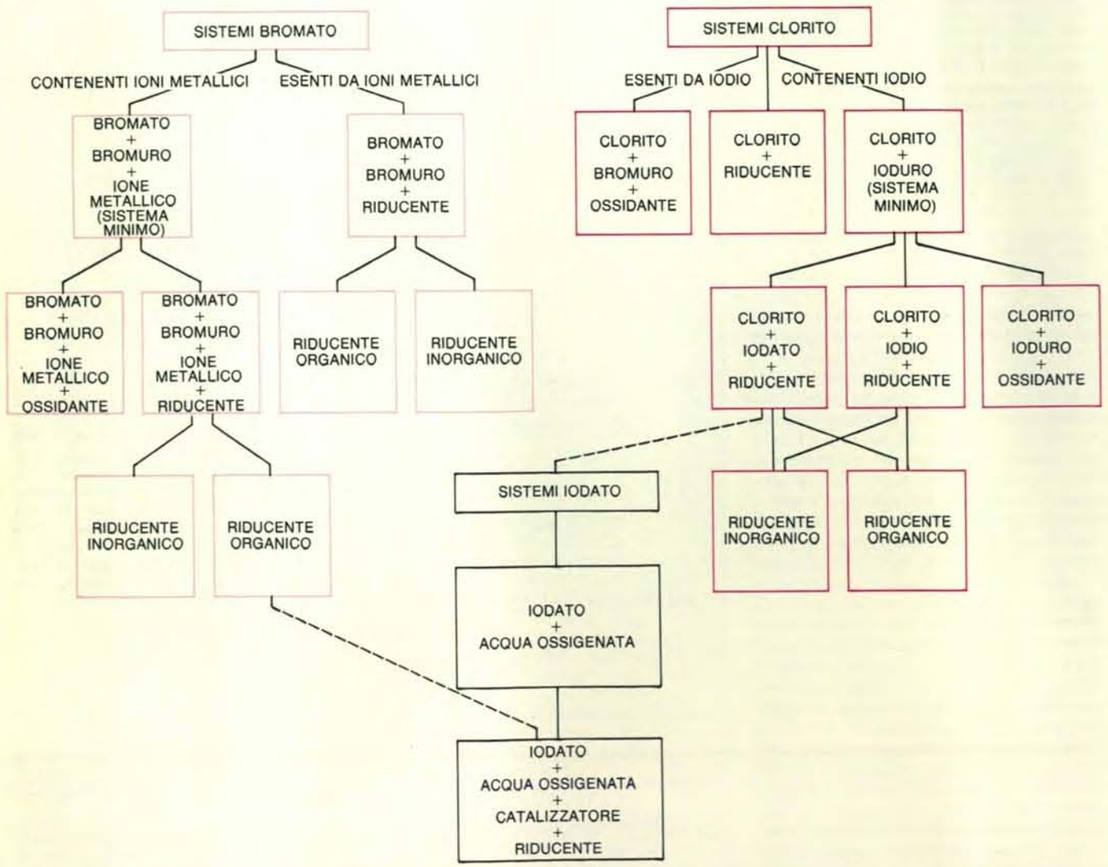
**PERSONAL COMPUTER**
di A. Gupta e H. D. Toong (n. 174)

Passando in rassegna hardware, software, applicazioni e diffusione di questi calcolatori, ci si rende conto che essi sono accessibili anche a persone non dotate di una preparazione tecnica specifica.

Il nostro piano per ideare un sistema chimico oscillante comprendeva i quattro stadi seguenti. Trovare un sistema autocatalitico, far avvenire la reazione in un reattore a flusso continuo, variare le condizioni fino a trovare una regione di bistabilità e, quindi, introdurre un'altra sostanza capace di influire in modo diverso sui due rami di bistabilità e pertanto di indurre oscillazioni. La nostra ricerca sulla letteratura chimica per le reazioni autocatalitiche ci condusse alla scoperta di molte reazioni interessanti. Due ci sembrarono particolarmente attraenti poiché avevano come intermedio comune lo iodio. Una reazione coinvolgeva gli ioni iodato e l'arsenito, la seconda gli ioni ioduro e il clorito. Come si potrà immaginare, una reazione poteva avere un'auspicabile influenza perturbante sull'altra. La reazione con iodato e arsenito si rivelò ben presto bistabile. Quando venne introdotto il clorito, il sistema composito iniziò quasi immediatamente a oscillare. Il primo sforzo sistematico per ideare un nuovo sistema chimico oscillante fu, quindi, coronato da successo.

Variando i componenti che entrano nel reattore a flusso continuo, noi abbiamo da tempo esteso il sistema oscillante iniziale clorito-iodato-arsenito a un gruppo di due dozzine di sistemi oscillanti correlati, che hanno tutti in comune lo ione clorito. Possiamo tracciare una classificazione dei sistemi oscillanti a base di clorito e stiamo cominciando a vedere in che modo possono essere messi in relazione con i sistemi oscillanti delle famiglie del bromato e dello iodato.

Benché i sistemi bromato siano stati fra i primi a essere scoperti e il loro meccanismo sia meglio conosciuto di quello dei sistemi clorito, la varietà dei sistemi clorito, che abbiamo trovato deliberatamente, è di gran lunga più ampia. Due fattori, uno chimico e l'altro storico, spiegano questo apparente paradosso. Dal punto di vista chimico, ogni famiglia di sistemi chimici oscillanti ha un membro minimo (più semplice) dal quale gli altri possono essere derivati per aggiunta di altre sostanze. Il più semplice sistema oscillante a base di clorito, clorito più ioduro, oscilla entro un'ampia gamma di condizioni. Al contrario, il più semplice sistema oscillante a base di bromato, bromato più bromuro più uno ione metallico, funziona solo entro limiti ben precisi. Infatti, anche dopo che l'esistenza del sistema oscillante più semplice a base di bromato era stata prevista da un ricercatore, Kedma Bar-Eli dell'Università di Tel Aviv, la sua ricerca non diede frutti finché due di noi (Orbán ed Epstein) riuscirono a trovarlo seguendo il nostro metodo sistematico basato sul diagramma a più fasi.

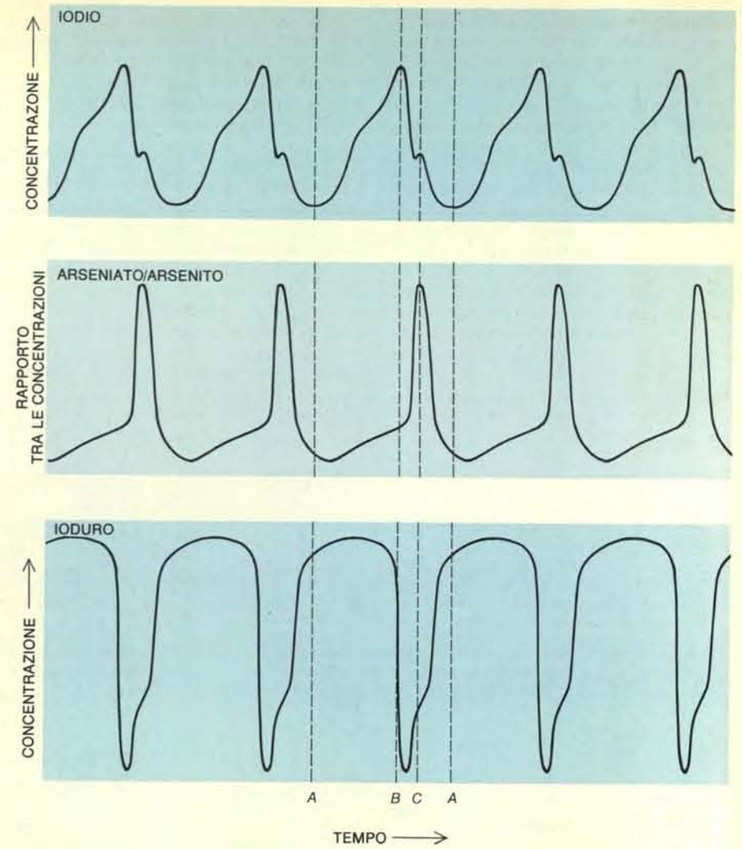



**I principali tipi di sistemi chimici oscillanti sono raggruppati, in un tentativo di classificazione, in tre famiglie principali. Il sistema originario di Belousov-Zhabotinsky e tutte le sue numerose varianti si trovano nel ramo della famiglia del bromato che ha ioni metallici. La sistemazione del sistema bromato nel ramo esente da ioni metallici è opera di Körös e di uno degli autori (Orbán). Tutti i sistemi oscillanti a base di clorito sono stati scoperti dagli autori alla Brandeis University. I sistemi al bromato che comportano riducenti inorganici sono stati messi a punto nella stessa sede in condizioni di flusso continuo. La famiglia dello iodato ha il minor numero di membri. Il primo sistema elencato in essa è la reazione oscillante illustrata nel 1921 da William C. Bray dell'Università della California a Berkeley e ignorata per 50 anni. Il secondo sistema nella famiglia dello iodato è un ibrido tra i sistemi Belousov-Zhabotinsky e Bray. Linee tratteggiate collegano i sistemi che hanno alcuni reagenti in comune. Un riducente è una sostanza che cede elettroni a un'altra; ossidante è una sostanza che li acquista.**



Dal punto di vista storico, il primo sistema oscillante a base di bromato fu scoperto prima che entrassero in funzione i reattori a flusso continuo. Sistemi del genere oscillavano facilmente in condizioni di flusso zero. I sistemi a base di clorito, come la grande maggioranza dei nuovi sistemi, mostrano un comportamento periodico solo in condizioni di flusso continuo, poiché il flusso è necessario a mantenerli lontani dall'equilibrio. Un sistema che oscilla in condizioni di flusso zero quasi sempre oscillerà anche in condizioni di flusso, ma raramente avviene il contrario. La maggior parte dei sistemi bromato fu originariamente ricercata in condizioni di flusso zero, mentre quelli clorito furono in origine ricercati solo in condizioni di flusso. Di conseguenza, la varietà di sistemi a base di clorito, che sono stati scoperti, ha presto superato la varietà dei sistemi a base di bromato. (Trovammo successivamente, però, che molti sistemi oscillanti a base di clorito operano anch'essi in condizioni di flusso zero.) Si può dire che l'estrema facilità di oscillazione dei primi sistemi a base di bromato ritardò la ricerca di altri sistemi. Soltanto recentemente i reattori a flusso continuo sono stati impiegati in modo intensivo nella ricerca di nuovi sistemi oscillanti a base di bromato e allo studio di una terza famiglia, finora la più piccola: i sistemi a base di iodato.

Uno degli scopi della progettazione di un nuovo sistema chimico oscillante fu quello di poter disporre di un sistema il cui meccanismo potesse essere decifrato e confrontato con quello delle reazioni oscillanti precedentemente esaminate. L'analisi della famiglia dei sistemi oscillanti a base di clorito è ora ben avviata. Abbiamo completato gli studi delle velocità di numerosi processi componenti e pertanto abbiamo numerose tessere del mosaico; per avere una descrizione completa dello schema oscillante basta ora spostare gruppi di tessere finché combacino tutte insieme nel modo giusto.



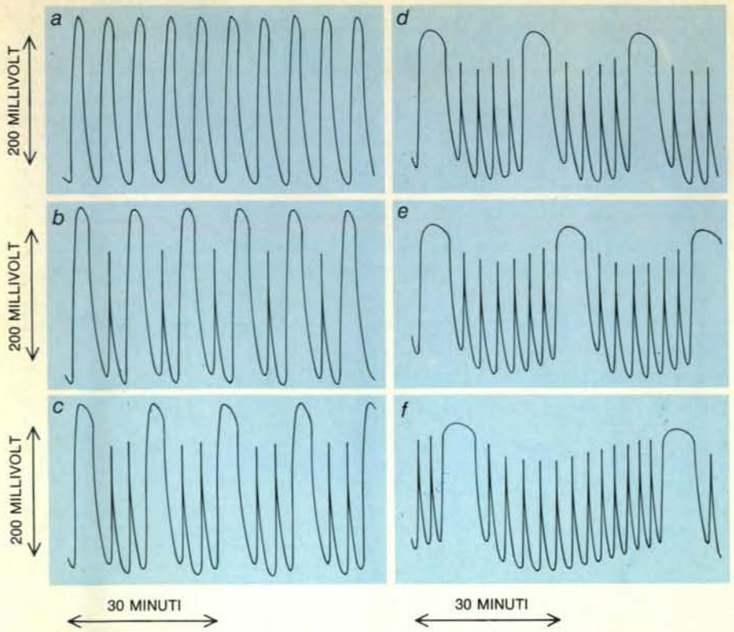

1 CLORITO + IODURO → CLORURO + IODIO
2 CLORITO + IODIO → CLORURO + IODATO
3 IODATO + IODURO → IODIO
4 IODIO + ARSENITO → IODURO + ARSENIATO

**L'andamento delle oscillazioni nel primo sistema oscillante messo a punto dagli autori può essere compreso qualitativamente considerando l'interazione tra quattro reazioni principali. Il sistema è alimentato con un flusso uniforme e bilanciato di tre soluzioni saline: una di clorito, la seconda di arsenito e l'ultima di iodato. All'interno del reattore a flusso continuo le concentrazioni dello iodio e dello iodato e il rapporto arseniato/arsenito oscillano, come appare dalle tre curve. Al punto *A* le concentrazioni di ioduro e di arsenito sono elevate, mentre quelle dello iodio e dell'arseniato sono basse, così che le reazioni predominanti sono la *1* e la *3*. Per effetto di queste reazioni aumenta bruscamente la concentrazione dello iodio, mentre incomincia a diminuire la concentrazione dello ioduro; il sistema raggiunge il punto *B* e si interrompono le reazioni *1* e *3*. Si svolgono allora le reazioni *2* e *4*, con l'esaurimento dello iodio e la conversione dell'arsenito in arseniato. Il sistema raggiunge il punto *C*, dove l'arsenito e lo ioduro si trovano ai loro livelli minimi, o quasi. L'arsenito viene reimmesso nel reattore e l'arseniato viene scaricato. La reazione *4* rigenera ioduro, mentre consuma iodio e pertanto il sistema ritorna in *A*.**

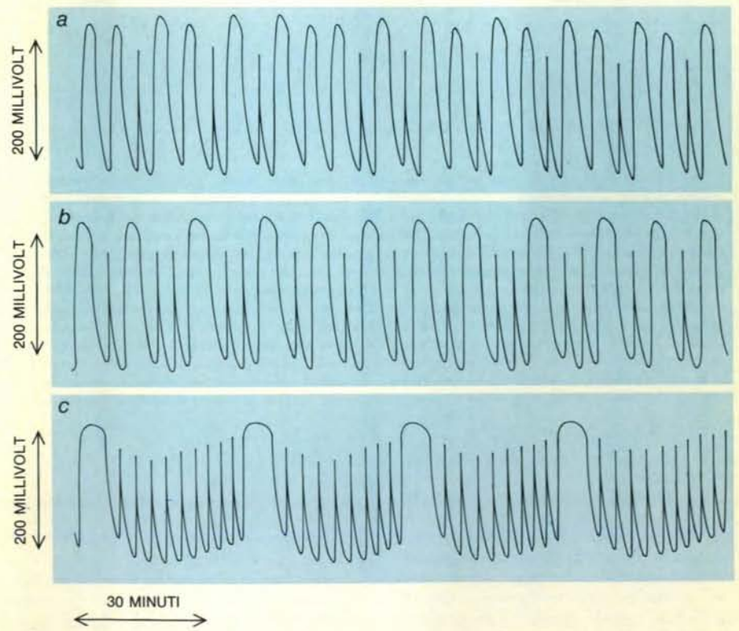
Siamo già in grado di dare un quadro schematico del modo in cui funziona il nostro primo sistema oscillante, quello costituito da soluzioni di sali di potassio o di sodio sotto forma di clorito, iodato e arsenito. Il sistema comporta solo quattro reazioni principali, ciascuna delle quali composta di numerosi stadi elementari. Due di essi comportano la reazione del clorito con lo ioduro o con lo iodio per dare nel primo caso cloruro e iodio e nel secondo cloruro e iodato. Nel terzo stadio lo iodato e lo ioduro reagiscono per formare iodio. Nel quarto lo iodio e l'arsenito formano ioduro e arseniato. Variazioni nelle velocità di queste quattro reazioni danno luogo a oscillazioni periodiche nelle concentrazioni di iodio e iodato e nel rapporto tra arseniato e arsenito *(si veda l'illustrazione in alto)*.

L'oscillazione chimica, benché affascinante di per se stessa, è connessa a numerosi altri fenomeni di almeno pari interesse. Uno è la formazione di strutture spaziali in un mezzo inizialmente omogeneo. Tali strutture si sviluppano, per esempio, in uno dei sistemi oscillanti che abbiamo scoperto, dove i reagenti sono clorito, ioduro e acido malonico. Quando è presente l'amido come indicatore, la soluzione ha inizialmente un colore uniforme rosso violaceo, a causa di un complesso costituito da ioduro, iodio e amido. Con il procedere della reazione appaiono punti bianchi, che si ingrossano fino a formare anelli e serie di anelli concentrici che si annientano quando urtano fra di loro. Un osservatore ha paragonato la comparsa dei punti bianchi nel color rosso circostante alla apparizione improvvisa di stelle. La comparsa di tali forme geometriche in una massa che prima era indifferenziata fa venire subito alla mente il genere di processi che permette alle cellule embrionali negli organismi animali di unirsi in tipi individuali destinati a diventare sangue, cervello o osso.

Un altro fenomeno affine, che ha fatto sorgere una grande controversia fra i




**Le oscillazioni periodiche nella reazione clorito-tiosolfato mostrano una varietà di modelli quando viene alterata la velocità di immissione in un reattore a flusso continuo. Ogni ciclo ha una oscillazione ampia e *n* oscillazioni piccole, dove *n* varia da 0 a 16. Ciò si verifica per tutti i cicli.**




**Un'oscillazione aperiodica, o caotica, si osserva talvolta in una stretta gamma di velocità di flusso, tra velocità che danno un comportamento periodico complesso. Una forma di caos consiste in un miscuglio casuale di due tipi di oscillazioni periodiche, una con *n* e l'altra con *n*+1 piccole oscillazioni per ogni ampia oscillazione. Una possibile spiegazione è che le variazioni incontrollate nella velocità di flusso facciano oscillare a caso il sistema da una velocità che dà oscillazione di tipo *n* a un'altra che dà oscillazione di tipo *n*+1. L'alternativa, favorita dagli autori, è che l'oscillazione caotica sia intrinseca nella dinamica della reazione chimica a determinate velocità di flusso.**

chimici, maggiore di quella sorta un quarto di secolo fa quando fu messa in dubbio la realtà delle reazioni oscillanti, è il «caos chimico». Il termine si riferisce al comportamento di una reazione in cui le concentrazioni non raggiungono valori costanti e neppure oscillano periodicamente, ma piuttosto aumentano e diminuiscono in un modo apparentemente casuale e imprevedibile. Noi abbiamo notato tali fluttuazioni nel sistema BZ e in uno dei nostri nuovi sistemi oscillanti a base di clorito, quando vengono fatti reagire in un reattore a flusso continuo sotto accurato controllo. Essi sono di particolare interesse per i matematici poiché se fossero per natura inerenti alla dinamica della reazione, essi rappresenterebbero esempi fisici di quegli oggetti matematici noti come «attrattori strani». Brevemente e in qualche modo grossolanamente, gli attrattori strani possono comparire quando alcune equazioni sono iterate, o risolte ripetutamente, con il risultato che una iterazione serve da entrata per la successiva. In un sistema di equazioni con un attrattore strano il percorso segnato da soluzioni successive sembra variare in modo imprevedibile da un ciclo all'altro.

Chimici scettici hanno ipotizzato che il caos chimico non rappresenti niente di più che fluttuazioni sperimentali incontrollate, per esempio nella temperatura o nella velocità di flusso, che spingono a caso un sistema da un modo di oscillazione periodica complessa a un altro. Sebbene un giudizio definitivo non sia stato ancora dato, un recente e accurato lavoro sperimentale e teorico di Harry L. Swinney e Jack Turner dell'Università del Texas a Austin, di Jean Claude Roux, Christian Vidal e collaboratori del Centre de Recherches Paul Pascal e di J. L. Hudson dell'Università della Virginia propone che il caos chimico sia un fenomeno genuino inerente alla dinamica di alcuni sistemi chimici oscillanti. Le implicazioni di questo caos intrinseco non sono ancora conosciute. Potrebbe darsi, come Henry Adams osservò una volta, che «il caos spesso genera la vita, quando l'ordine genera abitudine»?

Ora che i sistemi chimici oscillanti hanno dimostrato non solo di essere compatibili con le leggi della natura, ma anche di poter essere ideati e descritti sistematicamente, incominciano ad avere un ruolo sempre più importante in molti settori scientifici. Il passaggio da uno stato a un altro, che caratterizza una reazione oscillante, può serbare la chiave per capire i processi di regolazione nelle cellule viventi, come i meccanismi che innescano o arrestano la copiatura di un filamento di DNA o la contrazione di un muscolo. Le stesse forze che creano anelli e strati colorati in sistemi chimici oscillanti non sottoposti a rimescolamento possono anche essere responsabili della separazione degli anelli di Saturno e delle striature periodiche in alcune formazioni rocciose, spiegate con difficoltà dai processi geologici convenzionali.

Sembra inoltre plausibile che molte e

## AMBIENTE E INQUINAMENTO

LE SCIENZE *edizione italiana di* SCIENTIFIC AMERICAN *ha pubblicato su questo argomento numerosi articoli tra cui:*


**CIRCOLAZIONE GLOBALE DELL'INQUINAMENTO ATMOSFERICO**
di R. E.Newell (n. 32)

**INQUINAMENTO DA MERCURIO**
di L. J. Goldwater (n. 36)

**ENERGIA «PULITA» DA COMBUSTIBILI «SPORCHI»**
di A. M. Squires (n. 53)

**L'ELIMINAZIONE DEI RIFIUTI NELL'OCEANO**
di W. Bascom (n. 76)

**FARFALLE MELANICHE E ARIA PULITA**
di J. A. Bishop e L. M. Cook (n. 81)

**CATRAME PELAGICO**
di J. N. Butler (n. 86)

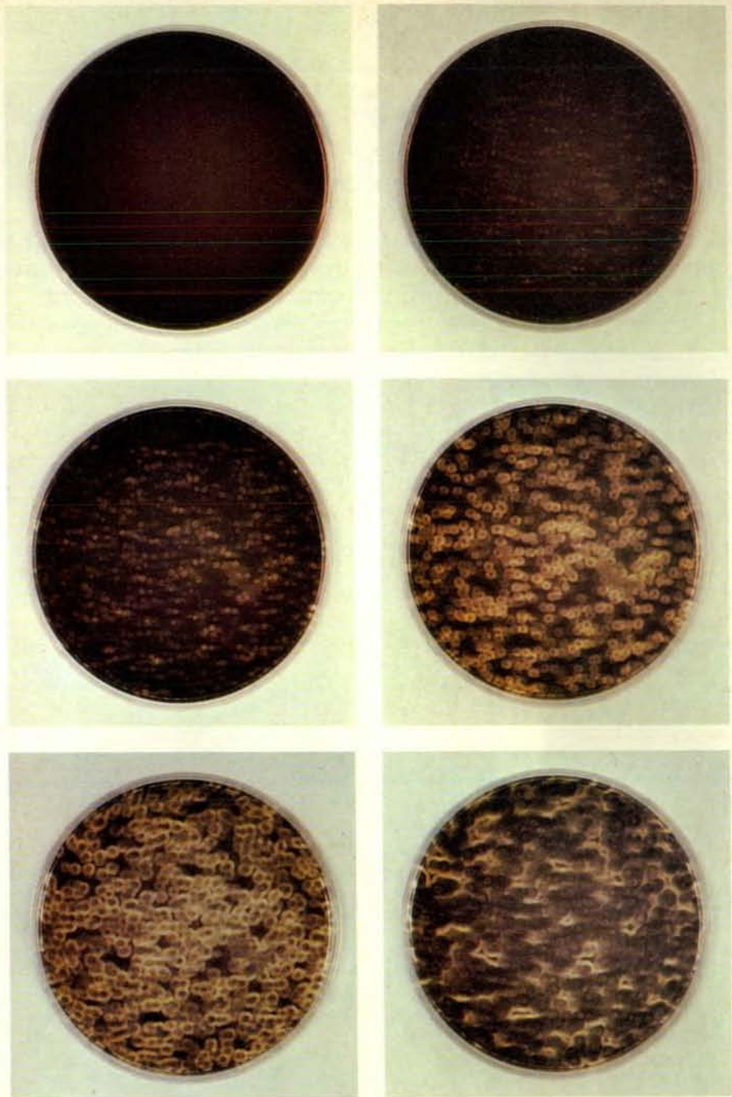
**L'INQUINAMENTO DA TRIZIO**
di G. Zamparo (n. 90)

**EFFETTI ANTROPOGENICI SULL'OZONOSFERA**
di F. Verniani (n. 102)

**IL CONFINAMENTO DELLE SCORIE RADIOATTIVE**
di B. L. Cohen (n. 110)

**IL PROBLEMA DELL'ANIDRIDE CARBONICA**
di G. M. Woodwell (n. 115)

**INQUINAMENTO DA PCB**
di H. Suzuki (n. 116)

**In un sistema chimico oscillante a base di clorito, ioduro e acido malonico, scoperto dagli autori, si sviluppa una struttura spaziale. La soluzione, posta in una piastra di vetro poco profonda, è inizialmente di un colore rosso brunastro uniforme per la formazione di un complesso tra iodio, ioduro e amido, usato come indicatore. Al procedere della reazione appaiono punti bianchi, che si ingrossano fino a formare anelli e serie di anelli concentrici, che si annullano quando si urtano. L'ultima fotografia è stata scattata circa 90 secondi dopo l'inizio della reazione.**

forse tutte le reazioni catalitiche che sono fondamentali per l'industria chimica procedano in modo oscillante. Il catalizzatore può alternarsi tra due forme in un intervallo di tempo troppo breve perché sia rilevato dai normali processi di analisi. Lo studio di reazioni oscillanti può fornire un terreno di prova per teorie sulla catalisi e, più in generale, sulla dinamica delle reazioni chimiche. Qualunque modello possa prevedere con successo il complesso comportamento di sistemi oscillanti dovrebbe essere in grado di trattare senza difficoltà i fenomeni più semplici osservati in reazioni meno strane. Aumentando il numero dei sistemi chimici oscillanti trovati o ideati deliberatamente, si può sperare di trovare la chiave per penetrare in una varietà di fenomeni, compresi quegli «orologi» e «caroselli» che operano all'interno del nostro corpo.

# Il DNA mitocondriale

*Il mitocondrio, un organello generatore di energia, possiede un proprio sistema genetico. L'organizzazione dei suoi geni e la versione del codice genetico «universale» che esso dà sono diverse da quelle di nucleo e batteri*

di Leslie A. Grivell

Un miliardo e mezzo o due miliardi di anni fa, quando sulla Terra c'era ancora poco ossigeno, un batterio primitivo che viveva in condizioni precarie, traendo energia dalla fermentazione anaerobica di molecole organiche, inglobò una cellula più piccola che aveva sviluppato in un certo modo la capacità di respirare. Questo evento costituì una svolta importante nella evoluzione organica. La respirazione libera molta più energia della fermentazione e la crescente abbondanza di ossigeno nell'atmosfera deve essere stata la forza trainante di questa relazione simbiotica, sviluppatasi tra le cellule, con la cellula aerobica che generava energia in cambio di protezione e nutrimento, offerti dall'ospite di dimensioni maggiori.

Col tempo, la cellula inglobata e altre come essa sarebbero divenute organelli subcellulari, trasmessi dalle cellule ospiti alla loro progenie. Alla fine, anche le cellule ospiti si sarebbero modificate, sviluppando altre strutture subcellulari e membrane interne e segregando il proprio materiale genetico in cromosomi all'interno di un nucleo. Queste cellule furono le antenate di tutte le attuali cellule eucariote (nucleate): protozoi, alghe e le singole cellule dei funghi, delle piante e degli animali. I discendenti attuali di quegli antichi batteri simbionti, dotati di respirazione, sono i mitocondri, le centrali energetiche della cellula eucariota.

I mitocondri sono organelli ovali o vermiformi, con diametro di circa mezzo micrometro e lunghezza da due a cinque micrometri. Questa è più o meno la dimensione di molti batteri attuali. Il mitocondrio possiede una membrana esterna e una interna, che presenta ampie ripiegature e racchiude una matrice liquida. Esso è la sede della fosforilazione ossidativa, la fonte primaria di energia cellulare. Nella matrice liquida, le molecole organiche derivate dalla degradazione degli alimenti vengono ossidate in una serie di reazioni chimiche, note come ciclo dell'acido citrico. Gli elettroni rimossi nel corso dell'ossidazione vengono trasportati lungo una catena di complessi enzimatici respiratori, allineati sulla membrana interna, i quali guidano la fosforilazione dell'adenosindifosfato in adenosintrifosfato (ATP), la moneta di scambio energetico universale delle cellule. Il citoplasma delle cellule eucariote (cioè la regione della cellula esterna al nucleo) contiene da pochi a molte centinaia di mitocondri: quanto maggiore è il fabbisogno energetico della cellula, tanto maggiore è il numero dei mitocondri.

Forse proprio perché discendono da un batterio libero, i mitocondri hanno un materiale genetico proprio e un proprio meccanismo per esprimerlo. Questa possibilità, che i mitocondri avessero i propri geni, distinti dal genoma contenuto nel nucleo cellulare, cominciò a farsi strada fin dal 1949, quando Boris Ephrussi trovò che la capacità del lievito da panificazione di realizzare una fosforilazione ossidativa sembrava controllata da qualche fattore presente nel citoplasma e non nel nucleo. Solo nel 1966, tuttavia, venne isolato e caratterizzato il primo DNA mitocondriale di vertebrato (pollo). Da quella data in poi, molti DNA mitocondriali di vari organismi sono stati oggetto di uno studio intensivo in numerosi laboratori europei e statunitensi. Affascinanti interrogativi furono: Che cosa fa il DNA mitocondriale? In che modo si organizza e si esprime? Perché la cellula dovrebbe mantenere un secondo sistema genetico oltre a quello contenuto nel nucleo? Cominciano a farsi strada oggi le risposte ai primi due quesiti e ciò permette perlomeno di avanzare ipotesi sulla terza e più ampia questione.

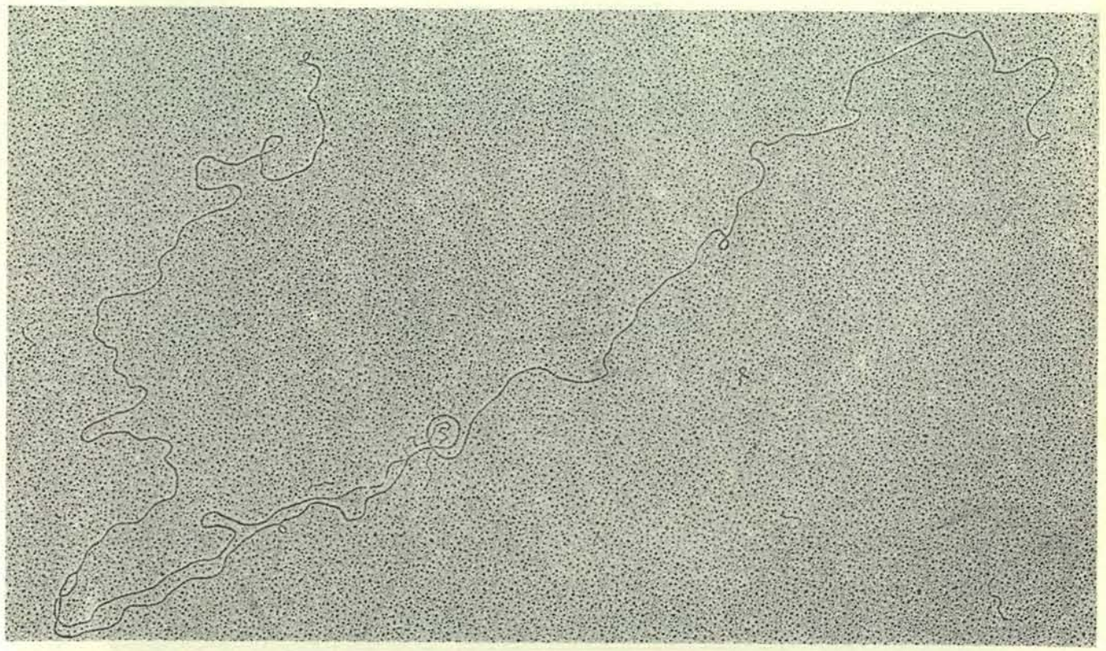
Negli ultimi due anni sono stati riportati nuovi e importanti dati, grazie soprattutto all'applicazione della genetica molecolare e di potenti nuove tecniche alla rapida determinazione della sequenza dei nucleotidi (le unità costruttive del DNA) nel genoma. La completa sequenza nucleotidica del DNA mitocondriale umano è stata pubblicata all'inizio del 1981 da B. G. Barrel e collaboratori del laboratorio di Frederick Sanger presso il Medical Research Council Laboratory of Molecular Biology di Cambridge; contemporaneamente Giuseppe Attardi del California Institute of Technology ha analizzato la trascrizione del DNA mitocondriale umano in RNA. Subito dopo, il gruppo di Barrell ha pubblicato la sequenza del DNA mitocondriale di bue e David A. Clayton e collaboratori del California Institute of Technology quella del DNA mitocondriale di topo. Nel frattempo, anche la sequenza di una grande parte del genoma mitocondriale di lievito è stata dipanata da Alexander Tzagolov della Columbia University e dal suo gruppo. Quella che si sta sempre più delineando è la diversità, per molti aspetti fondamentali, dei vari DNA mitocondriali nei riguardi di qualunque DNA noto. Le loro insolite caratteristiche potrebbero rappresentare le ultime tracce del primitivo organismo aerobico, che molto tempo fa colonizzò i batteri fermentanti, dando così inizio alla rivoluzione eucariota.

### *Il genoma e le sue proteine*

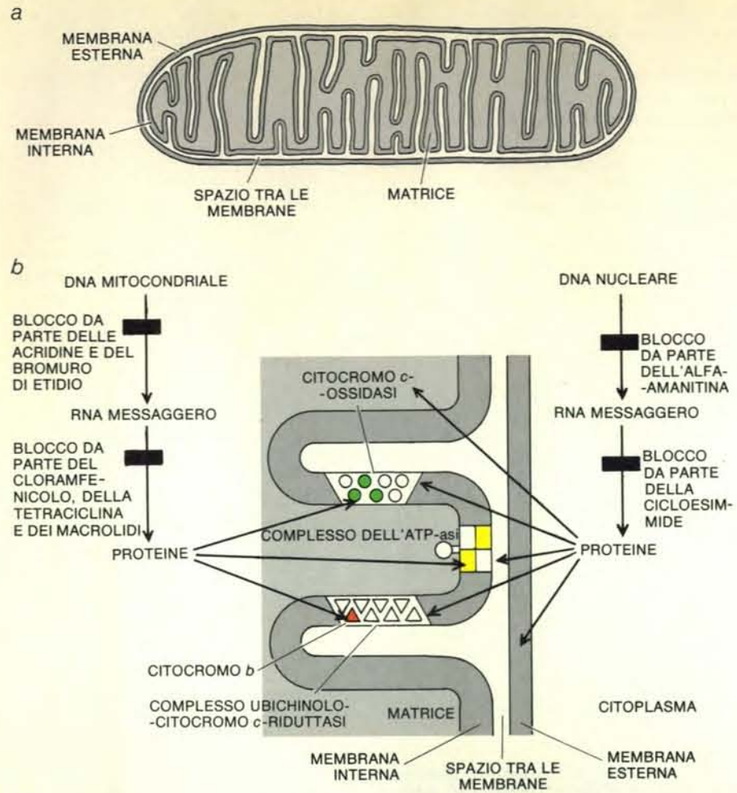
Il DNA mitocondriale è come quello dei batteri, nel senso che consiste di un'unica doppia elica di DNA nudo, mentre il DNA nucleare degli eucarioti forma un complesso ben aderente con le proteine ed è ripartito tra un certo numero di cromosomi distinti. Tranne che in alcuni protozoi, l'elica di DNA è conformata (di nuovo come nei batteri) ad anello. Nella maggior parte delle specie, tuttavia, il genoma mitocondriale è molto più piccolo di quello dei batteri più semplici: come dimensioni, è più vicino al genoma di molti virus. Le dimensioni variano ampiamente secondo l'organismo. I mitocondri degli animali hanno perloppiù circa 15 000 nucleotidi, mentre il DNA mitocondriale del lievito è cinque volte più lungo e quello delle piante ancora cinque volte di più. In buona parte la differenza di lunghezza tra DNA animale e DNA di lievito si può attribuire alla notevole diversità con cui è organizzata la stessa serie fondamentale di geni. Non si è ancora in grado di affermare se questo valga anche per il DNA mitocondriale delle piante.

Pur avendo il proprio materiale genetico, il mitocondrio non è però autosuffi-

**Il DNA mitocondriale del lievito di birra, *Saccharomyces carlsbergensis*, appare ingrandito di circa 60 000 diametri in questa microfotografia elettronica eseguita da Ernst F. J. Van Bruggen dell'Università statale di Groningen, in Olanda. La molecola appare come un doppio filamento ad anello di DNA, «avvolto a superelica», con una circonferenza di 26 micrometri. Essa è costituita da circa 75 000 nucleotidi.**



**Il genoma mitocondriale del protozoo *Tetrahymena pyriformis* si presenta come una molecola lineare. In questa microfotografia ripresa da Annika C. Arnberg e da Van Bruggen, esso appare ingrandito di circa 26 000 diametri e si trova in uno stadio precoce di duplicazione. Un centro di duplicazione si è costituito presso la metà della molecola e si sta estendendo mentre nuovo DNA viene sintetizzato lungo entrambi i filamenti.**




**La struttura e la funzione del mitocondrio vengono definite congiuntamente da geni presenti nel nucleo cellulare e da altri presenti nello stesso mitocondrio. Questo organello, che compare in sezione longitudinale *(a)*, ha una membrana esterna e una membrana interna profondamente ripiegata, che racchiude una matrice liquida. In tutti gli organismi esaminati finora, il DNA del mitocondrio codifica per gli RNA ribosomali e gli RNA di trasporto, presenti nell'organello stesso, e per elementi del meccanismo proteinosintetico. Altri geni mitocondriali codificano per le proteine. Nel lievito ne sono note appena sei, specificate dal genoma mitocondriale *(b)*, tre subunità dell'enzima citocromo *c*-ossidasi *(in verde)*, due subunità dell'enzima ATP-asi *(in giallo)* e la proteina citocromo *b (in rosso)*. Tutte e sei prendono parte alla fosforilazione ossidativa, la principale fonte di energia cellulare. La maggior parte delle proteine che costituiscono l'organello o che sono necessarie per il funzionamento del suo sistema genetico vengono codificate, invece, nel nucleo. Si può sapere se responsabile di una particolare sintesi è il sistema genetico nucleare o quello mitocondriale, bloccando chimicamente uno dei sistemi di trascrizione o di traduzione.**

ciente sotto l'aspetto genetico: infatti, la maggior parte delle proteine che sono responsabili della struttura e della funzione di questo organello sono codificate nel nucleo. Il DNA del mitocondrio specifica l'RNA dei suoi ribosomi (gli organelli su cui sono sintetizzate le proteine) e i suoi RNA di trasporto (che trasportano amminoacidi specifici sui ribosomi affinché siano riuniti e costituiscano proteine). Finora si è trovato che, come per le proteine, i DNA mitocondriali sintetizzano RNA messaggero che codifica solo per alcune subunità dei complessi enzimatici respiratori, situati nella membrana interna dell'organello. Il resto delle proteine della membrana interna, così come quelle della matrice (comprendenti gli enzimi che intervengono nel ciclo dell'acido citrico) e della membrana esterna, è tutto codificato dai geni nucleari, sintetizzato nel citoplasma dai ribosomi e quindi importato nel mitocondrio. Oltre a rifornire il mitocondrio della maggior parte dei suoi materiali strutturali, il nucleo codifica per tutti gli enzimi necessari alla replicazione del DNA mitocondriale e alla sua trascrizione in RNA; inoltre, il nucleo codifica per le proteine che si combinano con l'RNA per formare i ribosomi, e specifica altri fattori necessari alla sintesi delle proteine mitocondriali. In altre parole, il sistema genetico mitocondriale si conserva solamente grazie a un considerevole investimento da parte del nucleo e del meccanismo cellulare per la sintesi delle proteine.

Nella cellula vivente i contributi disparati dei due sistemi genetici sono coordinati, ma vi è una certa libertà d'azione in questo accoppiamento: la cellula può vivere per un certo tempo mentre l'attività genetica, o del nucleo o dei mitocondri, è bloccata. Per esempio, la sintesi proteica dei mitocondri è bloccata da antibiotici antibatterici come il cloramfenicolo, la tetraciclina e il gruppo dei macrolidi; non viene, invece, alterata dalla cicloesimmide, che inibisce l'attività dei ribosomi citoplasmatici. Si possono sistemare le cose in modo che, mentre un sistema è bloccato, le proteine sintetizzate siano marcate con un isotopo radioattivo, determinando così ciò che ogni sistema esattamente produce.

Questi esperimenti di marcatura hanno identificato i prodotti del DNA mitocondriale in un certo numero di organismi. Identificare i geni che codificano per tali prodotti e individuarne la posizione sull'anello di DNA è stato, però, un compito arduo, perlomeno fino a quando non sono stati messi a punto i nuovi metodi per ricostruire la sequenza del DNA. L'identificazione dei geni e la costruzione delle loro mappe mediante analisi genetica dipendono dalla possibilità di annotare e seguire le mutazioni che interferiscono con l'attività di un gene senza uccidere l'organismo. In assenza di ossigeno, un lievito può cavarsela con la sola fermentazione e riesce così a sopravvivere senza mitocondri funzionanti. È stato così possibile annotare e caratterizzare in esso mutazioni che sono letali in altri organismi; come risultato, il DNA mitocondriale del lievito è stato il primo a essere studiato in modo approfondito.

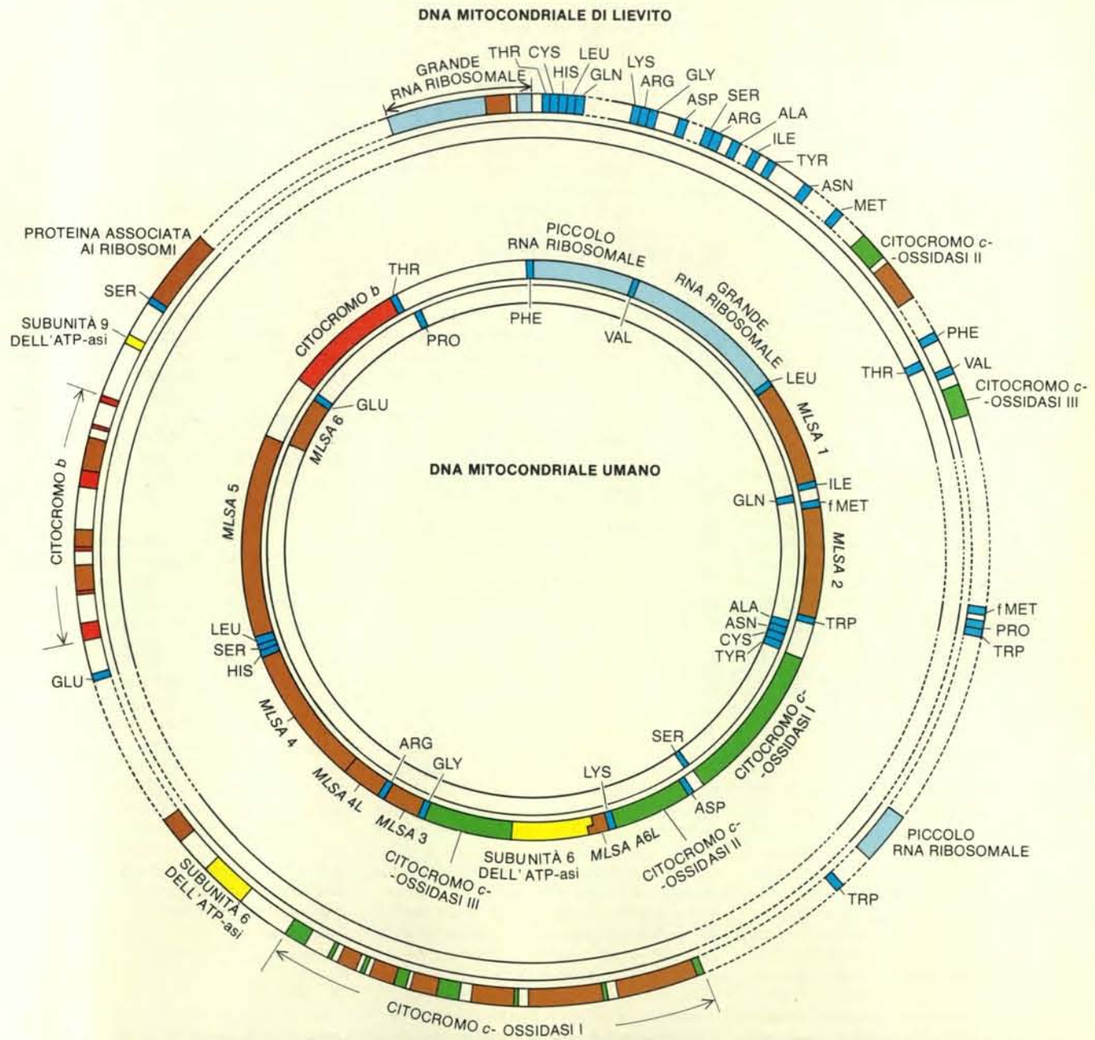
L'analisi genetica combinata con la tecnica di ricostruzione delle sequenze del DNA ha ormai fornito inventari di geni sia per i genomi mitocondriali umani sia per quelli di lievito, assieme a mappe particolareggiate della loro disposizione. I mitocondri del lievito e quelli umani hanno essenzialmente gli stessi geni, ma l'organizzazione di questi ultimi è sorprendentemente diversa.

### *L'organizzazione del genoma*

Ogni nucleotide del DNA è caratterizzato da una o l'altra di quattro basi e la sequenza nucleotidica dei due filamenti della doppia elica è complementare: l'adenina *(A)* su un filamento si appaia con la timina *(T)* sull'altro filamento, mentre la guanina *(G)* si appaia con la citosina *(C)*. Le triplette di basi su un filamento (il filamento codificante) servono da parole di codice (codoni), le quali, trascritte nell'RNA, specificano o un particolare amminoacido o un segnale per dare inizio o terminare la traduzione. Nel DNA che costituisce i cromosomi degli eucarioti, i codoni sono perloppiù una minoranza: le file di triplette che costituiscono un gene sono affiancate da spaziatori lunghi e non codificanti e da regioni che fungono da segnalazione, entrambi non trascritti in RNA, inoltre da altri segmenti che sono trascritti in RNA messaggero, ma non sono tradotti. Gli stessi geni sono spesso suddivisi in segmenti separati da sequenze intercalari non codificanti, chiamate «introni».

Il DNA mitocondriale umano, per contro, è un modello di economia. Ha solo 16 569 coppie di basi e sembra che praticamente ognuna di esse sia stata messa in servizio, con l'uno o l'altro filamento che serve di volta in volta da filamento codificante. In effetti, i geni sono così ravvicinati tra loro che possono esservi nel mezzo, se pure ve ne sono, solo pochi nucleotidi non codificanti. Vi sono geni per due RNA ribosomali, per 22 RNA di trasporto e per cinque proteine della membrana; si possono annoverare, inoltre, otto moduli di lettura senza assegnazione *(MLSA)*. (Si tratta di regioni in cui triplette di nucleotidi in successione possono essere lette con codoni che specificano amminoacidi e in cui la sequenza di tali triplette non viene interrotta da codoni che indicano la fine. Queste regioni devono specificare una proteina, ma questo prodotto proteico non è stato ancora identificato.) Nessuno di questi geni contiene introni. Pochissimo DNA codifica per l'RNA messaggero, che in seguito non viene tradotto in proteina.

Sia nei genomi batterici sia nei genomi eucarioti non mitocondriali, la trascrizione di ciascun gene è controllata individualmente da una interazione tra enzimi, che intervengono nella trascrizione, e sequenze segnale, presenti nelle regioni non codificanti del DNA. La scarsità di questo DNA nel genoma mitocondriale umano suggerisce che la trascrizione debba essere regolata in qualche modo insolito. In che modo tutto questo avvenga sta




**I mitocondri umani e quelli del lievito hanno essenzialmente la stessa serie di geni, mentre la loro organizzazione è diversa. I due DNA vengono raffigurati qui come doppi cerchi concentrici (un cerchio per ogni filamento della doppia elica), ma in scale diverse: la molecola del DNA di lievito è in effetti cinque volte più lunga di quella umana. L'intera sequenza nucleotidica del genoma umano è stata determinata da B. G. Barrell e collaboratori; è anche nota la maggior parte della sequenza del lievito *Saccharomyces cerevisiae*, di cui rimangono indefiniti alcuni segmenti *(linee tratteggiate)*. I segmenti in colore mostrano l'estensione dei geni per proteine note *(in verde, giallo e rosso)*, per gli RNA ribosomali *(blu chiaro)* e per gli RNA di trasporto *(blu intenso)*, ognuno dei quali è marcato con l'abbreviazione per l'amminoacido specifico; i segmenti in colore marrone indicano, infine, i moduli di lettura senza attribuzione *(MLSA)*, presumibilmente geni che codificano per proteine che non sono state ancora identificate. Il genoma umano è estremamente compatto e presenta scarso DNA non codificante intercalato tra i geni. Nel genoma del lievito, d'altra parte, vi sono lunghi segmenti non codificanti e parecchi geni sono interrotti da sequenze intercalari («introni»), alcune delle quali includono a loro volta dei moduli di lettura che non hanno ancora avuto alcuna attribuzione.**

cominciando a essere chiaro, come risultato degli studi accurati condotti da Attardi e collaboratori al California Institute of Technology. Sembra che su ogni filamento del DNA mitocondriale umano vi sia un unico importante promotore, o sito di inizio per la sintesi dell'RNA. Pertanto, il prodotto primario della trascrizione è una copia di RNA a tutta lunghezza di ogni filamento. I due prodotti primari della trascrizione vengono quindi scissi e danno gli RNA ribosomali, gli RNA di trasporto e gli RNA messaggeri che saranno in seguito tradotti in proteine. La scissione può aver luogo anche mentre l'RNA viene sintetizzato. In assenza di DNA non codificante, quali siti nei prodotti primari di trascrizione vengono riconosciuti dagli enzimi che operano la scissione? I prodotti di trascrizione che danno luogo a RNA di trasporto si trovano proprio a fianco di quasi ogni gene più lungo, e la scissione ha luogo esattamente all'inizio e alla fine di ogni sequenza che diventerà RNA di trasporto. Questo fa pensare che il prodotto primario della trascrizione si ripieghi in modo tale che la forma caratteristica degli RNA di trasporto fornisca i siti di riconoscimento per gli enzimi che intervengono sull'RNA.

Analogamente, la mancanza di uno spazio segnalatore attorno agli RNA messaggeri fa pensare a insolite strategie per il controllo della traduzione. Il segnale per l'inizio della traduzione è lo stesso del codone per l'amminoacido metionina. Gli RNA messaggeri non mitocondriali sono preceduti da sequenze che servono da siti di riconoscimento per i ribosomi, i quali si legano a esse e marcano così come segnale di «inizio» il successivo codone per la metionina. I ribosomi mitocondriali umani devono essere in grado in qualche modo di riconoscere certi codoni per la metionina come codoni di inizio, senza avvalersi di un precedente sito di riconoscimento.

Il problema di come arrestare la sintesi proteica alla fine del gene è più complesso. Negli RNA messaggeri non mitocondriali il termine della traduzione è segnalato da uno di parecchi codoni terminatori di catena, tra i quali vi è *UUA* (*U* sta per uracile, la base dell'RNA analoga alla timina). La maggior parte dei messaggeri mitocondriali umani o mancano di un codone terminatore di catena (continuando così nel gene successivo e sovrapponendosi ad esso; questo gene è di solito un gene per un RNA di trasporto) oppure hanno un codone terminatore di catena che viene troncato dalla scissione (mediata dagli RNA di trasporto), descritta sopra. I geni citati terminano tutti o con *U* o con *UA*. In che modo, allora, si forma il codone terminatore? L'RNA messaggero mitocondriale umano, come la maggior parte degli altri messaggeri eucarioti è «poliadenilato», vale a dire possiede una lunga coda di nucleotidi dell'adenina, subito dopo il punto di scissione. Questa coda poli-*A* offre proprio quello che il meccanismo di traduzione richiede: un segnale di arresto *UUA*.

Il DNA mitocondriale di lievito è tanto prodigo quanto la sua controparte umana è parsimoniosa. Il genoma del lievito non è stato ancora completamente ricostruito come sequenza, ma sembra che questo DNA codifichi per due RNA ribosomali, circa 25 RNA di trasporto, sei proteine di membrana e più di dieci proteine non identificate. Nel lievito (in modo specifico nel lievito da panificazione *Saccharomyces cerevisiae*), i geni sono dispersi su un DNA cinque volte più lungo. Le sequenze codificanti sono separate da lunghi tratti di DNA, che consistono quasi interamente di nucleotidi delle basi *A* e *T*. Almeno tre geni sono suddivisi in segmenti da introni. Lunghe sequenze all'inizio e alla fine dei geni sono trascritte, ma non tradotte.

### *Il codice non è universale*

Ogni codone dell'RNA messaggero viene «letto» dall'anticodone di un RNA di trasporto, che trasporta appunto un amminoacido specifico da collocare in una catena proteica. Il codice genetico che correla ogni codone del DNA (o dell'RNA messaggero suo prodotto di trascrizione) con un particolare amminoacido è esattamente lo stesso nei virus, nei batteri e negli animali, dai vermi all'uomo. Venne pertanto ritenuto universale, ma in realtà non è così: i mitocondri ne hanno una propria versione, che è soggetta a variazione nei diversi organismi. I mitocondri dei mammiferi leggono *AGA* e *AGG* come segnali di arresto e non come codoni per l'amminoacido arginina e traducono *AUA* come metionina e non come isoleucina; *AUA* e *AUU* possono servire come segnali di inizio al posto di *AUG*. D'altra parte, nei mitocondri delle muffe *Neurospora* e *Aspergillus*, *AGA*, *AGG* e *AUA* vengono tradotti secondo il codice «universale». Nel lievito, come nei mammiferi, *AUA* codifica per la metionina, mentre i quattro codoni che cominciano con *CU* vengono letti come treonina invece che come leucina. Nei mitocondri dei mammiferi, del lievito, di *Neurospora* e di *Aspergillus*, l'amminoacido triptofano viene specificato, oltre che dal normale *UGG*, da *UGA*, mentre nel granturco il codone supplementare per il triptofano sembra essere *CGG*.

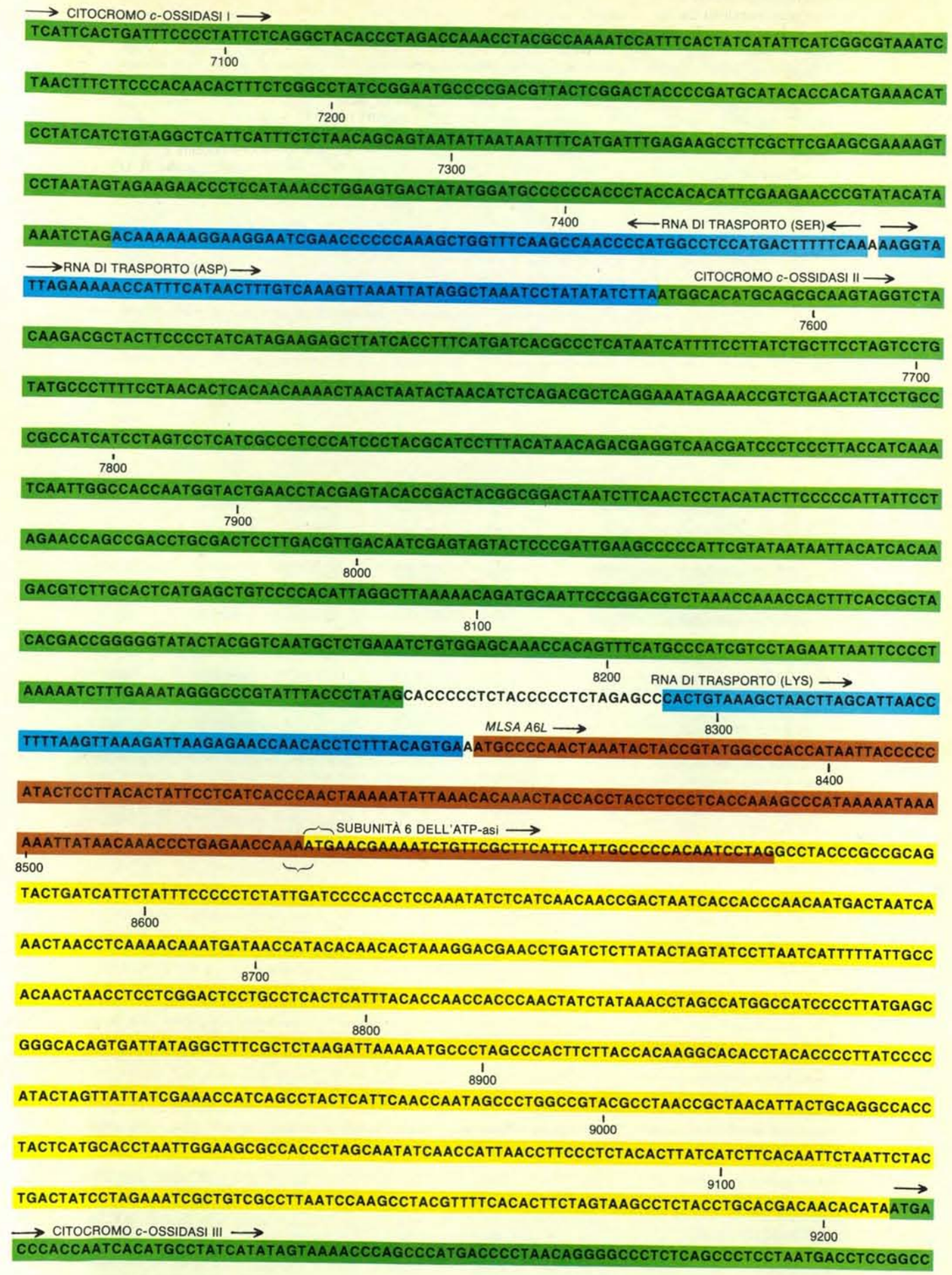
Esistono 61 codoni (cioè $4^3$, meno tre codoni terminatori di catena), ma gli organismi se la cavano con un minor numero di RNA di trasporto perché un solo RNA di trasporto è in grado di leggere più di un codone. E può farlo perché c'è una certa libertà (ossia un certo «ondeggiamento») nell'appaiamento della terza base di un codone con certe basi nella prima posizione di un anticodone. Secondo le regole dedotte alcuni anni fa da F. H. C. Crick, 32 RNA di trasporto (ma non meno) dovrebbero essere in grado di leggere i 61 codoni. Eppure l'analisi delle sequenze mostra che il DNA mitocondriale umano codifica per soli 22 RNA di trasporto, il DNA mitocondriale di lievito per 25. Non sembra che altri RNA di trasporto siano importati dal citoplasma, eppure 61 codoni vengono letti. Come?

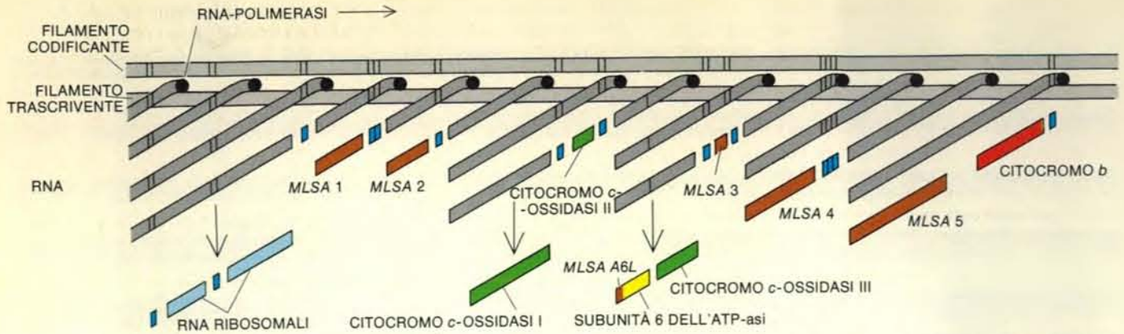
Sembra che ciascuno dei 22 RNA di trasporto dei mitocondri di mammifero sia in grado di leggere una «famiglia» di due o quattro codoni sinonimi *(si veda l'illustrazione a pagina 110)*. Gli RNA di trasporto per famiglie di due codoni avrebbero degli anticodoni ondeggianti convenzionali *G*:*U*, cioè degli anticodoni che cominciano con *G*, una base in grado di appaiarsi con *U* o con *C*, oppure che cominciano con *U*, una base che può appaiarsi con *G* o con *A*. Gli RNA di trasporto per famiglie di quattro codoni ignorano, invece, le regole di Crick e leggono - a quanto pare - tutti e quattro i codoni sinonimi. I loro anticodoni hanno una *U* nella posizione 5′. O questa *U* si appaia in un certo modo con tutte e quattro le basi del codone 3′ oppure gli anticodoni leggono solo le prime due basi e ignorano l'ultima. La possibilità di questo processo di lettura «due su tre» è stata suggerita parecchi anni fa da Ulf Lagerkvist della Facoltà di medicina dell'Università di Göteborg.

Un interessante interrogativo è come alcuni RNA di trasporto con la base *U* in posizione ondeggiante possano leggere - a quanto si sostiene - due codoni al posto di quattro. Che cosa li trattiene dal riconoscere i codoni che terminano con *U* o *C*, e così quelli che terminano con *G* o *A*? Nei mitocondri del lievito e di *Neurospora* e in molti RNA di trasporto non mitocondriali un simile riconoscimento è reso impossibile da una modificazione chimica dell'*U* ondeggiante, ma finora l'analisi delle sequenze degli RNA di trasporto mitocondriali animali non ha rivelato alcuna modificazione in tale posizione. La maggior parte di questi RNA di trasporto possiede, però, una struttura non convenzionale ed è forse proprio questa che influenza la specificità dell'anticodone.

**Un campione della sequenza nucleotidica del DNA mitocondriale umano si ricava dai dati raccolti da Barrell e collaboratori. Compaiono qui solo 2200 nucleotidi dell'intera molecola, costituita da ben 16 569 nucleotidi. Questo è sufficiente, tuttavia, a evidenziare come i geni siano disposti in modo compatto. In una regione (che comincia con il nucleotide numero 8527), due geni sono in realtà sovrapposti, cioè gli stessi nucleotidi svolgono un doppio compito, codificando la fine del modulo di lettura senza attribuzione *(MLSA)* che viene definita *A6L*, e l'inizio del gene per la subunità 6 dell'ATP-asi. (Il modulo di lettura è diverso, come viene indicato dalle due graffe.) La sequenza è quella del filamento che funge da codificante per la maggior parte dei geni illustrati e ha quindi la stessa sequenza dei corrispondenti RNA messaggeri e di trasporto. L'RNA di trasporto per l'amminoacido serina è, in realtà, specificato dal filamento complementare, che viene trascritto nella direzione opposta, come indicato dalla corrispondente freccia nella figura.**








**Il DNA mitocondriale umano viene trascritto in RNA e simultaneamente questo è trasformato per dar luogo agli RNA ribosomali, di trasporto e messaggeri maturi, secondo quanto afferma Giuseppe Attardi del California Institute of Technology. Lo schema indica in che modo viene trascritto il principale filamento codificante, partendo da un unico importante sito promotore *(a sinistra)*, per dar luogo a un unico prodotto di trascrizione. Questo RNA viene poi tagliato in corrispondenza delle estremità delle sequenze per gli RNA di trasporto *(rettangoli)*, che fiancheggiano praticamente ogni gene. Il prodotto di trascrizione iniziale e i suoi segmenti in seguito trasformati sono indicati in grigio. Gli RNA maturi sono invece indicati nei colori corrispondenti a quelli degli schemi delle tre precedenti illustrazioni.**

**I codoni terminatori di catena, che segnalano la fine della traduzione in proteina, mancano in sette prodotti di trascrizione mitocondriali umani. In alcuni la sequenza si imbatte in una sequenza codificante che si trova a valle; in altri il terminatore che consta di tre nucleotidi, viene tagliato, e quindi troncato, quando l'RNA viene sottoposto al processo di trasformazione. Il segnale necessario viene fornito in modo non convenzionale dalla poliadenilazione, un evento normale nella trasformazione dell'RNA, in cui una fila di adeninnucleotidi *(A)* viene aggiunta a ogni prodotto di trascrizione per ottenerne un RNA messaggero. In questo caso gli *A* ampliano il codone terminatore *UAA*.**

*Gli introni del lievito*

Geni suddivisi in segmenti separati (geni discontinui) furono scoperti per la prima volta nel 1977 ed è chiaro ormai che molti geni nei cromosomi dei mammiferi, degli uccelli e degli anfibi (come pure alcuni geni di organismi eucarioti inferiori) sono di questo tipo, con sequenze intercalari non codificanti che interrompono gli «esoni» codificanti per una proteina. Quando un gene di questo tipo si esprime, viene innanzitutto trascritto, introni e tutto il resto, in un RNA messaggero precursore. Questo RNA viene quindi «rielaborato», cioè gli introni vengono tagliati ed eliminati e gli esoni ricongiunti, o saldati (l'operazione viene detta con termine inglese *splicing*), così da costituire un RNA messaggero maturo, che viene tradotto in una proteina. Gli introni dei geni cromosomici, per quanto se ne sa, non codificano affatto.

Nei mitocondri del lievito, le cose stanno in modo diverso. Parecchi anni fa, Piotr P. Slonimsky e collaboratori del Centre de Génétique Moléculaire di Gif-sur-Yvette, in Francia, trovarono che mutazioni a carico di un introne del gene mitocondriale del lievito per il citocromo *b* (uno degli enzimi respiratori) influivano sullo *splicing* dell'RNA trascritto da quel gene. Una serie di eleganti esperimenti genetici, accoppiati con l'analisi delle sequenze del DNA, ne hanno rivelato la ragione: l'eliminazione del secondo introne del gene è mediata da una proteina che viene in parte specificata da un segmento dell'introne stesso!

Varie forme del gene per il citocromo *b* sono state studiate ormai in differenti ceppi di lievito. La forma più complessa ha sei esoni (designati con le sigle da *B*1 a *B*6) e cinque introni. Tre introni includono lunghi moduli di lettura che non hanno ancora ricevuto un'attribuzione. L'intero gene si esprime in un lungo prodotto primario di trascrizione. Quindi viene eliminato l'introne numero 1. L'effetto di questa fase consiste nell'unire l'esone *B*1 all'esone *B*2 e, nello stesso tempo, nel rendere accessibile il modulo di lettura dell'introne numero 2, cioè nel mettere le sue triplette in fase con quelle dei due precedenti esoni.

Il prodotto di trascrizione di *B*1, di *B*2 e della porzione dell'introne numero 2 che costituisce un modulo di lettura funge ora da RNA messaggero: viene tradotto in una proteina la cui testa ammino-terminale comprende i primi 143 amminoacidi del citocromo *b* (codificati da *B*1 e *B*2) e la cui coda carbossilico-terminale è codificata, invece, dal modulo di lettura dell'introne. Questa strana proteina aiuta a catalizzare il taglio e l'eliminazione dell'introne numero 2 e, nel far questo, distrugge il proprio RNA messaggero e quindi limita il proprio livello nel mitocondrio. Slonimski chiama questo fenomeno «meccanismo omeostatico basato sullo splicing», e definisce maturasi la proteina codificata dalla combinazione esone-introne, per indicare che può non essere di per sé un enzima in grado di catalizzare questo processo di resezione e saldatura, ma piuttosto un fattore che modifica in un certo modo la specificità di un preesistente enzima avente tale funzione, probabilmente un enzima codificato da un gene nucleare.

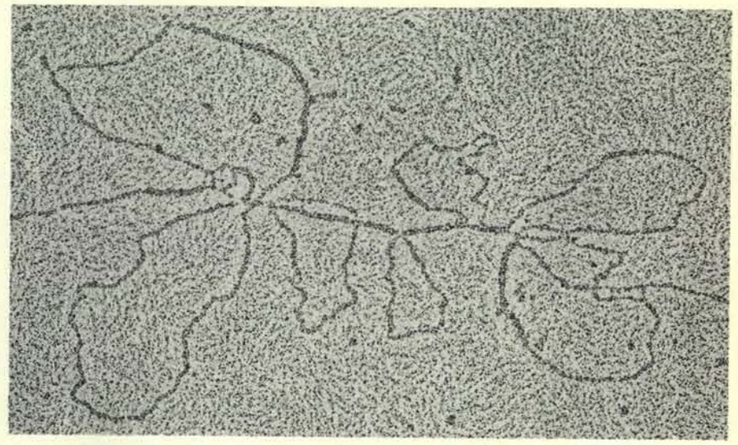
*Piccoli RNA circolari*

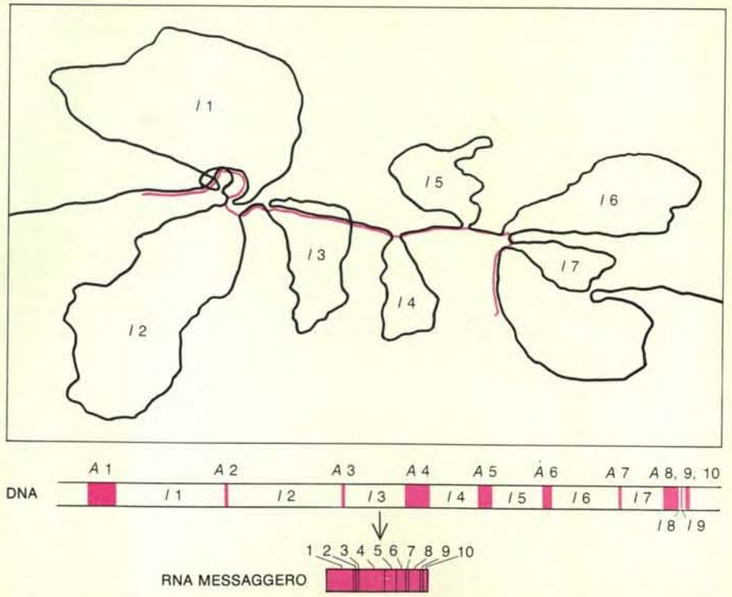
La presenza di una RNA-maturasi spiega in modo inequivocabile il comportamento di numerosi mitocondri mutanti che presentano lesioni all'interno dell'introne numero 2. In essi i prodotti di trascrizione del gene per il citocromo *b* non possono venir rielaborati, e quindi si accumulano lunghi RNA da cui è stato staccato solo il primo introne. Essi generano grandi quantità di proteine che danno una reazione crociata con gli anticorpi contro il citocromo *b*, ma che non sono citocromo *b*. La sintesi di tali proteine viene bloccata se nell'esone *B*1 sono introdotte mutazioni che fanno terminare la sintesi proteica. Esse devono essere allora delle maturasi difettose, che derivano dalla traduzione degli esoni *B*1 e *B*2 e del lungo modulo di lettura che è presente nell'introne numero 2. La loro produzione eccessiva è il risultato diretto della loro incapacità di distruggere l'RNA messaggero che ha diretto la loro sintesi.

Il meccanismo omeostatico basato sullo splicing non è probabilmente limitato al secondo introne del gene per il citocromo *b*. Gli introni numero 3 e 4 di quel gene e i primi quattro introni del gene per la subunità I della citocromo *c*-ossidasi (un'altra proteina della membrana interna) incorporano anch'essi lunghi moduli di lettura, resi accessibili, e vi sono prove che, nell'intervento su questi, siano interessate maturasi specifiche per gli introni.

Nei mitocondri, alcuni eventi di resezione e saldatura generano piccoli RNA ad anello insoliti, scoperti da Annika C. Arnberg dell'Università statale di Groningen, quando collaborava con Gertjan van Ommen e con me in studi di microscopia elettronica, condotti allo scopo di riuscire a costruire mappe dei prodotti di trascrizione mitocondriali. Risultò che tali anelli erano introni staccati da RNA precursori, e che in un secondo tempo avevano formato anelli chiusi, a quanto pare grazie alla formazione di un legame chimico covalente. Essi sono stati associati oggi con i tre introni del gene per la subunità I della citocromo *c*-ossidasi e con un introne (l'introne numero 2, di cui si è già parlato) del citocromo *b*. L'anello associato al citocromo *b* e un altro associato a un introne con un lungo modulo di lettura, che non ha ancora avuto una attribuzione, sono staccati - a quanto pare - da maturasi codificate nei mitocondri, mentre il distacco degli altri due anelli può dipendere da enzimi codificati nel nucleo.

Una volta Crick ha avanzato l'ipotesi che gli anelli di RNA potessero formarsi nel processo di resezione e saldatura nel caso che un enzima fosse in grado di unire le due estremità di un singolo introne alla stessa stregua di come poteva unire due esoni. Finora, laddove il processo di resezione e saldatura dà origine a RNA circolari, è stata riportata solo l'una o l'altra di queste due circostanze. Nel nucleo del protozoo *Tetrahymena*, Thomas R. Cech dell'Università del Colorado, a Boulder, ha riferito alcuni anni fa assieme ai suoi collaboratori l'esistenza di un introne ad






**Mediante microscopia elettronica è possibile ricostruire la mappa di questo gene discontinuo del DNA mitocondriale del lievito. La microfotografia, eseguita da Arnberg e Van Bruggen, e la mappa interpretativa sottostante mostrano una molecola ibrida: un singolo filamento di DNA, che incorpora il gene per la subunità I della citocromo *c*-ossidasi *(linea in nero sulla mappa)*, è stato fatto ibridare con l'RNA messaggero *(in colore)*, trascritto da esso. Nel DNA, le sequenze che codificano per la proteina sono interrotte dagli introni. Le sequenze complementari nel DNA e nell'RNA si sono appaiate e formano regioni a duplice filamento; gli introni, non avendo controparte nell'RNA, formano anse a filamento singolo. Quando un gene di questo tipo si esprime, l'RNA prodotto primario della trascrizione viene «reciso e nuovamente saldato» per rimuovere gli introni e lasciare soltanto gli «esoni», o regioni che codificano per la proteina. Le informazioni che si ricavano dagli esperimenti di ibridazione sono state combinate con i dati ricavati dalle sequenze e raccolti da Alexander Tzagoloff della Columbia University e da Lambert A. M. Hensgens, che lavora nel laboratorio dell'autore, per rivelare l'organizzazione degli esoni (da *A*1 ad *A*10) e degli introni (da *I*1 a *I* 9) e dell'RNA messaggero maturo dopo rielaborazione *(in basso)*. (Gli introni *I*8 e *I*9 sono troppo piccoli per distinguersi in una microfotografia.)**

| PRIMA LETTERA | SECONDA LETTERA | | | | TERZA LETTERA |
|---|---|---|---|---|---|
| | U | C | A | G | |
| U | PHE | SER | TYR | CYS | U |
| | PHE | SER | TYR | CYS | C |
| | LEU | SER | *STOP* | *(STOP)* TRP | A |
| | LEU | SER | *STOP* | TRP | G |
| C | LEU | PRO | HIS | ARG | U |
| | LEU | PRO | HIS | ARG | C |
| | LEU | PRO | GLN | ARG | A |
| | LEU | PRO | GLN | ARG | G |
| A | ILE (MET) | THR | ASN | SER | U |
| | ILE | THR | ASN | SER | C |
| | (ILE) MET | THR | LYS | (ARG) *STOP* | A |
| | MET | THR | LYS | (ARG) *STOP* | G |
| G | VAL | ALA | ASP | GLY | U |
| | VAL | ALA | ASP | GLY | C |
| | VAL | ALA | GLU | GLY | A |
| | VAL | ALA | GLU | GLY | G |

**Il codice genetico, in cui una tripletta di nucleotidi (designati dalle loro iniziali *U, C, A* e *G*) funge da codone, specificando un particolare amminoacido o un segnale di arresto (stop), veniva in passato ritenuto universale, mentre in realtà è diverso nei sistemi genetici dei mitocondri. La tabella qui sopra mostra in che modo le attribuzioni dei codoni deviino nei mitocondri umani da quelle del codice «universale»; ogni deviazione mitocondriale viene indicata in colore, con la traduzione normale tra parentesi. Ciascuno dei 22 RNA di trasporto presenti nei mitocondri umani «legge» o due codoni sinonimi *(caselle in grigio)* oppure quattro *(caselle in colore)*. La reazione degli RNA di trasporto della metionina nei confronti del codone non è ancora chiara.**

anello, staccato dal prodotto di trascrizione del gene per un RNA ribosomale. Quando tale gruppo cercò di identificare l'enzima che sembrava mediare questa resezione e circolarizzazione, giunse alla sorprendente conclusione che non esiste alcun enzima: sembra che la resezione e la circolarizzazione dipendano da proprietà intrinseche dello stesso RNA. L'RNA di *Tetrahymena* è l'unico esempio finora conosciuto di quello che Cech chiama un ribozima: cioè un RNA in grado di tagliarsi e saldarsi da sé. Ulteriori studi sugli anelli, presenti nei mitocondri e che non richiedono la presenza di maturasi per il distacco, dovrebbero riuscire a rivelare se anche nel loro caso sono interessati dei ribozimi.

*Perché gli introni?*

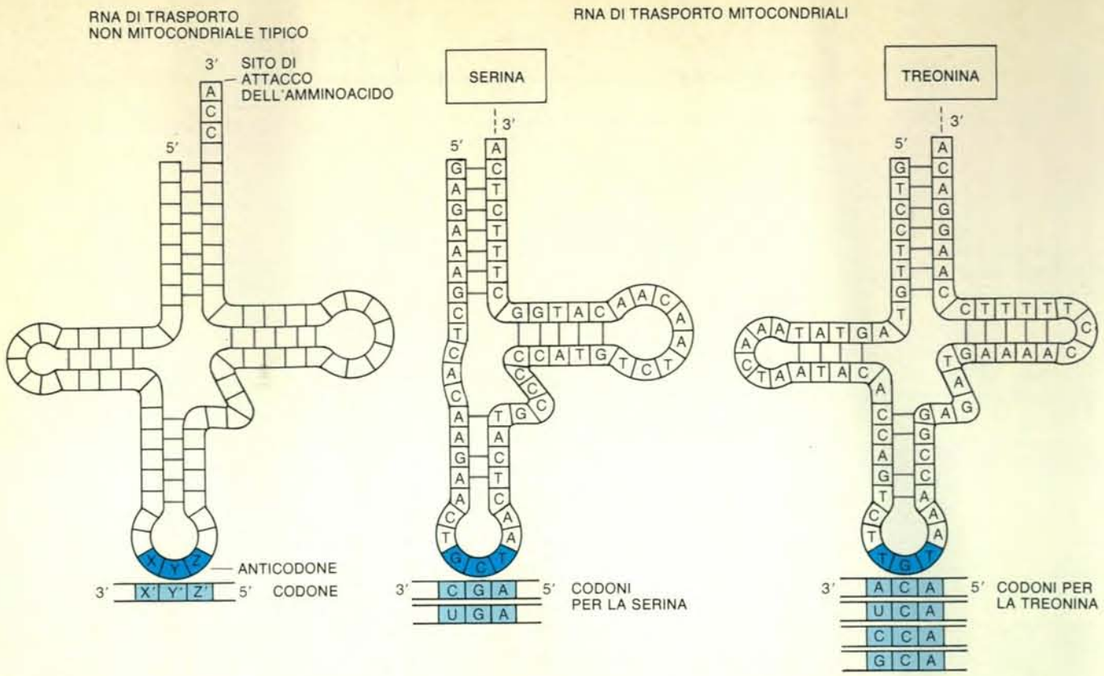
Una notevole proprietà della maturasi associata con l'introne numero 2 è la sua non essenzialità. Ceppi di lievito con versioni più brevi del gene per il citocromo *b* sono comuni e tali geni mancano spesso dei primi tre introni; gli esoni, quindi, da *B*1 a *B*4 risultano fusi. L'informazione originariamente presente negli introni mancanti non si trova altrove nel genoma mitocondriale del lievito, il che fa pensare che la maturasi sia l'equivalente mitocondriale del peggior tipo di burocrate: una persona che esiste solo per occuparsi della propria esistenza. In realtà, i tre introni mancanti sono membri di una classe di introni mitocondriali del lievito, definiti «opzionali» da Johan Sanders mentre si trovava nel nostro laboratorio presso l'Università di Amsterdam, alcuni anni orsono. Altri esempi si possono rilevare nei geni per il maggiore dei due RNA ribosomali e per la subunità I della citocromo *c*-ossidasi. Per classificare come opzionale un introne basta semplicemente trovare un ceppo che ne sia privo e che sia tuttavia funzionale. Finora è stato esaminato solo un numero limitato di ceppi di lievito, ma può darsi che, alla fine, tutti gli introni mitocondriali di lievito risultino opzionali. La natura opzionale di anche solo pochi introni pone in una nuova luce alcune questioni fondamentali: Perché alcuni geni sono discontinui e altri no? Perché lo stesso gene esisterebbe sia nella forma continua sia nella forma discontinua? Da dove derivano gli introni opzionali, e gli introni in generale?

Walter Gilbert della Harvard University ha sostenuto che, nei genomi nucleari e virali, gli introni possono conferire un vantaggio selettivo, promuovendo una evoluzione più rapida. Aumentando notevolmente la lunghezza di un gene, essi aumentano la probabilità di vari eventi di ricombinazione in cui si interscambiano segmenti analoghi di due alleli, o versioni alternative dello stesso gene. Una ricombinazione del genere potrebbe riunire in un unico allele due mutazioni vantaggiose, presenti ciascuna in un diverso allele, così da creare un gene doppiamente vantaggioso. Potrebbe anche creare geni completamente nuovi grazie a quello che Gilbert chiama «il rimescolamento degli esoni», cioè riunendo come esoni di un singolo gene parecchie sequenze codificanti, che in precedenza avevano specificato differenti proteine. Inoltre, una mutazione in una singola base in corrispondenza del limite tra un introne e un esone può modificare la modalità di resezione e saldatura e, pertanto, generare RNA messaggeri, che codificano per proteine molto diverse. I cambiamenti indotti da tali mutazioni nella resezione e saldatura non devono necessariamente essere assoluti, e quindi il DNA colpito può diventare un gene «con molteplici scelte»: può codificare sia per la proteina originaria sia per quella nuova, a seconda degli eventi di resezione e saldatura che si pongono in alternativa, il che permette all'organismo di provare una soluzione, senza cancellare dalla sua memoria la vecchia soluzione.

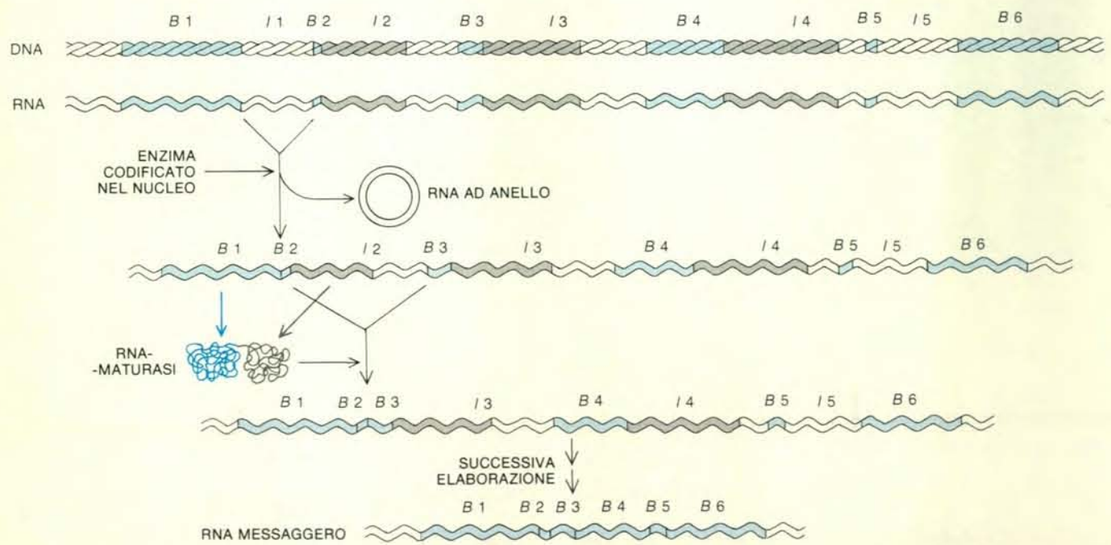
In che modo compaiono gli introni? L'introduzione di sequenze estranee in un gene essenziale è di solito letale. Anche se un meccanismo di elaborazione dell'RNA elimina tali sequenze dal prodotto primario di trascrizione di un gene, perché forze selettive dovrebbero conservare un introne appena introdotto? La cellula non trae alcun vantaggio immediato dalla presenza di una sequenza intercalare e certamente non può prevedere che essa le sarà utile in avvenire. Queste considerazioni hanno indotto W. Ford Doolittle III, della Dalhousie University nella Nuova Scozia, a proporre che gli introni non siano mai stati introdotti nelle cellule eucariote, ma facciano piuttosto parte della loro eredità evolutiva.

Secondo Doolittle, gli attuali procarioti (cioè i batteri), come pure il nucleo delle cellule eucariote, sono probabilmente derivati da un organismo ancestrale primitivo in grado di evolversi rapidamente in quanto possedeva un meccanismo di duplicazione e trascrizione del DNA, oltre che di traduzione, poco preciso e non fedele. L'informazione genetica di un simile organismo avrebbe dovuto essere codificata in modo ridondante perché una parte qualsiasi di essa potesse essere espressa in modo corretto. Una organizzazione fatta di geni frammentati potrebbe essere stata essenziale per assicurare che i prodotti di trascrizione degli esoni venissero almeno occasionalmente riuniti per dirigere la sintesi di proteine funzionali. Sempre secondo Doolittle, a mano a mano che il meccanismo genetico diventava più accurato, una simile precauzione diventò meno necessaria; la duplicazione e la trascrizione del DNA ridondante e non informazionale divennero sempre più onerose e ci fu una pressione per eliminarle. Il vantaggio di questa eliminazione fu una maggiore efficienza così come viene esemplificata dai geni semplificati dei moderni batteri e dei virus che li infettano. Il prezzo pagato è stata la perdita di una certa potenzialità di ulteriore cambiamento evolutivo.

Che dire degli introni nei geni mitocondriali? Molte delle proposte avanzate per spiegare gli introni dei geni nucleari non sono soddisfacenti per i geni mitocondriali, non solo perché non spiegano i moduli di lettura che non hanno ancora ricevuto una attribuzione, ma anche perché le principali previsioni non sono comprovate dai fatti. La ritenzione passiva di introni con moduli di lettura senza attri-




**Gli RNA di trasporto hanno in genere una struttura stereotipata a trifoglio *(a sinistra)*, il che deriva da un appaiamento *(linee in grigio)* tra basi nucleotidiche complementari. (Nell'RNA *A* si appaia con *U* e *G* con *C*. Negli RNA di trasporto mitocondriali illustrati qui, la sequenza è quella del DNA più che dell'RNA e così si nota che *A* si appaia con *T*.) Molti RNA di trasporto mitocondriali hanno, inoltre, una forma aberrante: quello per la serina *(al centro)* manca di un intero braccio. Un anticodone di tre nucleotidi su ciascun RNA di trasporto legge un codone dell'RNA messaggero. In alcuni casi, un RNA di trasporto può leggere più di un solo codone perché, nell'appaiamento tra la prima base del suo anticodone e la terza base di un codone vi è una certa libertà, o «ondeggiamento». La *X* dell'anticodone a sinistra potrebbe appaiarsi con una base diversa da *X'*. Molti RNA di trasporto nei mitocondri umani possono leggere - a quanto pare - fino a quattro codoni *(a destra)*.**




**La RNA-maturasi, un enzima specificato in parte dal secondo introne del gene mitocondriale del lievito per il citocromo *b*, aiuta a catalizzare il taglio dell'introne che la codifica. Il gene possiede sei esoni (da *B*1 a *B*6) e cinque introni. (Gli esoni e gli introni non sono rappresentati in questo grafico alla stessa scala.) Esso viene trascritto in un RNA precursore, da cui il primo introne è rapidamente eliminato mediante un enzima, codificato nel nucleo, e in grado di recidere e saldare il DNA. (Si accumulano - fatto sorprendente - copie dell'introne sotto forma di anelli a filamento singolo.) Il taglio finisce per unire gli esoni *B*1 e *B*2 a un lungo modulo di lettura all'interno dell'introne *I*2 *(in grigio)*, modulo che non ha ancora ricevuto una attribuzione. L'RNA così riunito funge da RNA messaggero, che specifica la maturasi, necessaria per il taglio di *I*2. La maturasi controlla in questo modo la propria sintesi. Vi sono prove che altre maturasi contribuiscano al taglio di *I*3 e di *I*4.**



buzione non è una buona spiegazione, perché il costo per la cellula che deve mantenere ed esprimere i numerosi geni necessari per staccare gli introni è notevole e un complesso sistema di questo genere deve essere molto vulnerabile alla demolizione per mutazione. La ritenzione per facilitare la ricombinazione è una spiegazione improbabile, dato che non esistono indicazioni che i nuovi geni mitocondriali siano sorti di recente in seguito al rimescolamento degli esoni o che proteine codificate nei mitocondri si stiano evolvendo nei funghi in modo più rapido delle loro controparti nei mitocondri dei mammiferi, i cui geni non hanno introni. Inoltre, anche se in teoria il rimescolamento degli esoni è un modo eccellente per diffondere mutazioni che conferiscono caratteri vantaggiosi (ad esempio, la resistenza a inibitori naturali della funzione mitocondriale), non vi è alcuna ragione per credere che tali mutazioni si fissino più efficacemente nei geni discontinui che nei geni continui. Infine, non vi è alcuna indicazione (anche se esistono dati validi qua e là) che qualunque gene discontinuo nei mitocondri di lievito sia un gene dalle molteplici scelte.

### *Gli introni e la repressione*

Gli introni con moduli di lettura che non hanno avuto alcuna attribuzione sono stati forse conservati, allora, nel DNA mitocondriale dei funghi, perché conferiscono un certo tipo di vantaggio selettivo. Quale potrebbe essere tale vantaggio? Una spiegazione plausibile, fornita da Piet Borst dell'Università di Amsterdam, è che introni di questo tipo possono svolgere un ruolo nella regolazione fine dell'espressione dei geni mitocondriali. Le cellule di lievito possono reprimere selettivamente la sintesi di certe proteine mitocondriali, quando crescono per fermentazione in assenza di ossigeno, e una parte di tale repressione può realizzarsi a livello della trasformazione dell'RNA. La proposta avanzata da Borst ha il vantaggio di spiegare perché la repressione sia selettiva: geni come quelli per il citocromo *b* e per la subunità I della citocromo *c*-ossidasi, che possiedono parecchi introni, il cui distacco è probabilmente sotto il controllo di una maturasi, dovrebbero essere particolarmente sensibili alla repressione ed è proprio questo che è stato trovato. L'argomento della repressione non spiega, però, l'esistenza di tutti gli introni mitocondriali del lievito. Non bisognerebbe lasciarsi sfuggire l'eventualità che almeno alcuni di quelli privi di moduli di lettura senza attribuzione potrebbero non avere affatto una funzione. Dato un meccanismo efficace per rimuoverli dall'RNA, potrebbe esservi una scarsa pressione selettiva per eliminarli dal DNA. I mitocondri potrebbero, in effetti, non avere il congegno necessario per svolgere quest'operazione in modo efficace. Gli introni potrebbero essere una forma di «DNA egoista», che perpetua se stesso per la propria salvezza.

Il fatto che alcuni introni nei mitocondri di lievito, di *Aspergillus* e e di *Neurospora* si trovino, nei geni che scindono, in siti strettamente corrispondenti fa pensare che gli antenati comuni dei geni possano aver avuto una organizzazione analoga. Dunque, perlomeno questi introni possono riflettere direttamente l'organizzazione a geni frammentati, propria del primitivo organismo endosimbiotico, da cui si pensa che si siano evoluti i mitocondri. Tuttavia, non tutti gli introni possono essere così vecchi. Confronti minuziosi di sequenze di introni nei lieviti, realizzati da François Michel e da Bernard Dujon del Centre de Génétique Moléculaire e da Linda Bonen dell'Università di Amsterdam rivelano molte similitudini sia nella struttura secondaria dell'RNA sia nelle sequenze che codificano per le proteine. Queste similitudini inducono con molta forza a credere che la maggior parte degli introni attuali e i loro moduli di lettura senza attribuzione siano membri di una famiglia multigenica, che ha avuto la sua origine in un unico gene ancestrale e si è poi espansa, nel corso dell'evoluzione, mediante duplicazione e trasposizione.

Questa duplicazione con mobilità è caratteristica del comportamento dei trasposoni, frammenti di DNA che, nei batteri e negli organismi superiori, si duplicano e inseriscono le loro copie in corrispondenza di siti casuali di un genoma. Gli introni molto simili del lievito hanno probabilmente avuto origine da essi. Se è avvenuto così, ciò spiegherebbe l'origine delle maturasi. La maggior parte dei trasposoni batterici codifica per una trasposasi, enzima che catalizza l'inserzione del trasposone. La trasposizione comporta una resezione esattamente in corrispondenza delle due estremità del trasposone, proprio come la resezione e saldatura dell'RNA esige tagli netti in corrispondenza delle estremità di un introne. Le maturasi potrebbero, dunque, essersi evolute dalle trasposasi.

### *Perché DNA mitocondriale?*

Perché un sistema genetico separato dovrebbe servire solo per un organello cellulare in mezzo a molti? Il sistema mitocondriale non può venir spiegato come una forma di amplificazione genica, necessaria per rifornire il mitocondrio di forti quantità di certe proteine. Elevati livelli di sintesi proteica si possono raggiungere con altri mezzi e, se l'amplificazione si rendesse necessaria, potrebbe anch'essa aver luogo nel nucleo. Né il DNA mitocondriale può venir spiegato come la fonte indispensabile di proteine che sono troppo insolubili per essere trasportate in corrispondenza del loro sito nella membrana interna. La subunità 9 del complesso ATP-asi, una delle proteine più idrofobe specificate nel lievito dal DNA mitocondriale, è specificata da un gene nucleare sia in *Neurospora* sia nei mammiferi, quindi sintetizzata sui ribosomi presenti nel citoplasma e, infine, importata nei mitocondri.

Se un sistema genetico mitocondriale separato non è necessario, perché esiste?


## I POLIMERI

**LE SCIENZE** *edizione italiana di* **SCIENTIFIC AMERICAN**

*ha pubblicato alcuni articoli su aspetti particolari della chimica delle macromolecole:*

**POLIMERI INORGANICI**
di H. R. Allcock (n. 70)
*I polimeri inorganici, ancora in fase di studio, presentano caratteristiche insolite e diverse da quelle dei polimeri organici impiegati nell'industria.*

◇

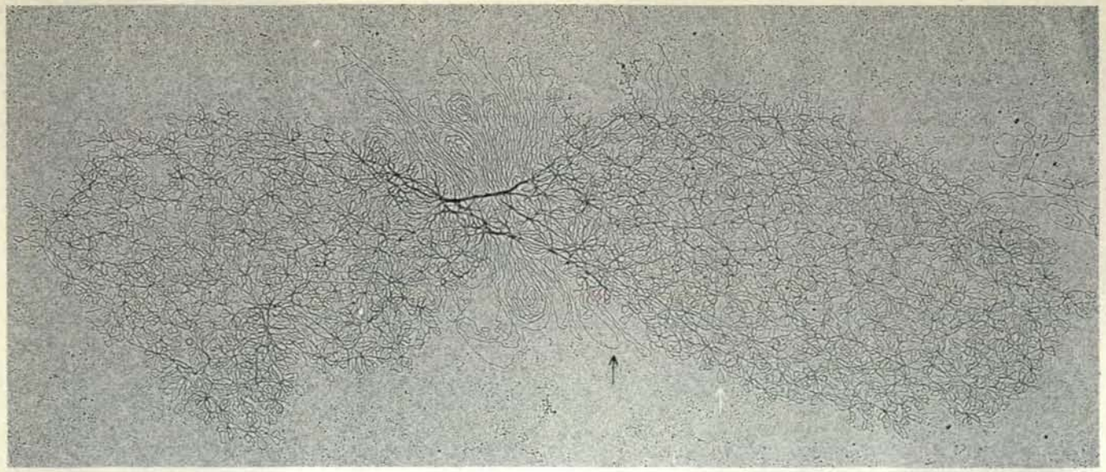
**LA MICROSTRUTTURA DEI MATERIALI POLIMERICI**
di D. R. Huhlmann e A. G. Kolbeck (n. 92)
*I progressi nella produzione di polimeri sintetici dipenderanno sia da una migliore conoscenza della loro complessità chimica, sia dalla realizzazione di tecniche più raffinate per controllare il loro ordinamento molecolare.*

◇

**I SILANI ORGANO-FUNZIONALI**
di G. Pantini (n. 113)
*Questi derivati del silicio, aumentando l'adesione fra matrici organiche di tipo polimerico e cariche inorganiche, migliorano notevolmente le proprietà meccaniche di vari materiali compositi.*

◇

**I POLIMERI FUNZIONALIZZATI**
di S. Leghissa (n. 140)
*Rappresentano il secondo aspetto della chimica macromolecolare dopo e accanto a quello più noto delle materie plastiche. I loro impieghi si irradiano su campi diversi e di alto interesse per la nostra vita futura.*



**Il DNA mitocondriale insolito del protozoo *Trypanosoma brucei* appare ingrandito di circa 16 000 diametri in questa microfotografia elettronica eseguita da J. H. Hoeimakers dell'Università di Amsterdam. Il DNA è in fase di segregazione: dopo essersi duplicato, si divide formando due genomi che si separano quando l'organello stesso si divide. Si presenta come un groviglio di maxianelli *(freccia in nero)* e minianelli *(freccia in negativo)*. Sembra che i maxianelli siano equivalenti al DNA mitocondriale di altri organismi; la funzione dei minianelli non è ancora nota.**

Borst ritiene che il DNA mitocondriale sia semplicemente un vicolo cieco evolutivo, ciò che rimane dell'ipotetico batterio simbiotico. Secondo la sua ipotesi, nei primi stadi dell'evoluzione esistevano scarse barriere al trasferimento dell'informazione tra il sistema genetico dell'endosimbionte e quello dell'ospite; quindi gradatamente la maggior parte dei geni dell'endosimbionte divennero parte integrante del genoma della cellula ospite. A mano a mano che la cellula eucariota divenne più sofisticata (cioè a mano a mano che il suo DNA si organizzò in cromosomi e fu segregato in un nucleo), e forse i codici genetici nucleare e mitocondriale si differenziarono l'uno dall'altro, la probabilità del trasferimento diminuì e, alla fine, divenne impossibile. Pochi geni rimasero, dunque, isolati nel mitocondrio e, poiché sono essenziali, persistono.

L'ipotesi attrae per la sua semplicità. Essa spiega perfettamente il fatto che un certo numero di enzimi mitocondriali codificati nel nucleo abbiano somiglianze con corrispondenti enzimi presenti nei batteri. Essa prevede che alcuni geni si troveranno nel DNA mitocondriale di un organismo e nel DNA nucleare di un altro, come nel caso del gene per la subunità 9 dell'ATP-asi. Contempla anche situazioni intermedie in cui le versioni sia mitocondriale sia nucleare dello stesso gene sono presenti nella stessa cellula. Ciò può valere per *Neurospora*, in cui Paul van den Boogaart ed Etienne Agsteribbe dell'Università statale di Groningen hanno trovato copie del gene per la subunità 9 dell'ATP-asi sia nel DNA mitocondriale sia in quello nucleare.

L'ipotesi di Borst prevede inoltre che alcuni DNA mitocondriali possano avere un numero di geni maggiore di quello che si trova nella serie di base presente sia nei mitocondri di lievito sia in quelli di mammifero. I mitocondri delle piante superiori hanno genomi estremamente cospicui, che vanno da circa 250 000 nucleotidi nel granturco ad anche 2,4 milioni nella famiglia delle cucurbitacee. Buona parte della differenza può essere dovuta alla presenza di DNA non codificante, ma esperimenti di laboratorio mostrano che i mitocondri delle piante sintetizzano più proteine delle loro controparti presenti nei funghi o nei mammiferi. Vi è, inoltre, una valida testimonianza che, nelle piante, il fenomeno della sterilità maschile (incapacità di produrre polline fertile) può essere il risultato di cambiamenti a carico del DNA mitocondriale.

### *L'evoluzione verso la diversità*

Mentre i principali prodotti dei geni mitocondriali sembrano essersi in gran parte conservati nel corso dell'evoluzione, lo stesso non vale per l'organizzazione e il modo di esprimersi di quei geni. La diversità dell'organizzazione viene esemplificata nel migliore dei modi nel DNA mitocondriale di *Trypanosoma*, un genere di protozoi unicellullari che include parassiti dell'uomo e degli animali domestici. Il DNA è un enorme groviglio di anelli legati (forse diecimila). Vi sono dei maxianelli, la cui lunghezza varia tra 20 000 e 36 000 nucleotidi. Hanno regioni omologhe con frammenti di DNA di lievito e vengono trascritti. Sono probabilmente la controparte del DNA mitocondriale, presente in altri organismi. La maggior parte del DNA, però, consiste di minianelli, la cui lunghezza è inferiore ai 3000 nucleotidi. Questi minianelli non vengono trascritti e hanno subito l'evoluzione a una velocità elevata, tipica del DNA non codificante. La funzione del vasto groviglio di minianelli non è nota, né è affatto chiaro il perché il DNA mitocondriale del tripanosoma dovrebbe avere una così insolita struttura.

Questa escursione finale a uno degli avamposti della biologia convincerà il lettore (lo spero) che i sistemi genetici mitocondriali sono diversi. La loro varietà è il risultato di una elevata velocità di evoluzione della sequenza nucleotidica, forse superiore di dieci volte alla velocità di evoluzione per i geni contenuti nel nucleo. La chiara mancanza di limitazioni all'evoluzione delle sequenze può essere messa in rapporto con la semplicità del sistema. La sua funzione è in primo luogo quella di sintetizzare un limitato numero di proteine, in quantitativi piuttosto costanti e potrebbe essere anche in grado di favorire mutazioni che sarebbero deleterie in un sistema adattato alla regolazione differenziale di molti geni. È anche possibile che in ciascun organismo differenti pressioni selettive abbiano sospinto i DNA mitocondriali verso soluzioni diverse, e talvolta estreme, del problema della espressione genica.

Il trovare che cosa sono queste pressioni potrebbe essere uno dei compiti per il futuro. Un altro compito potrebbe essere l'imparare, attraverso lo studio di questo sistema limitato, quali sono esattamente le richieste minime per un sistema genetico funzionale. Una miglior comprensione del genoma mitocondriale dovrebbe rivelare alcune delle vie evolutive che portano ai meccanismi di espressione genica nei batteri e negli eucarioti del mondo contemporaneo. A giudicare da quello che lo studio dei mitocondri ha già rivelato, le lezioni da imparare circa l'organizzazione e l'espressione del DNA mitocondriale dovrebbero avere applicazioni che si estendono ben oltre la membrana limite di questo modesto organello.

# TEMI METAMAGICI

*di Douglas R. Hofstadter*

## *La ricorsività nel Lisp, il linguaggio dell'intelligenza artificiale*

Ho terminato l'articolo del mese scorso con una notizia recentissima sul porscopino glazunkiano e ho pensato, quindi, che il miglior modo per iniziare l'articolo di questo mese fosse quello di aggiungere qualcosa su questo animale poco noto ma decisamente interessante. Come ricorderete, gli aculei di ciascun porscopino (tranne quelli più piccoli) sono porscopini di dimensioni minori. Il più piccolo dei porscopini non ha aculei ma un naso, e un naso molto importante dal momento che i glazunkiani basano il loro sistema monetario su tale piccolo oggetto. Prendiamo in considerazione il valore dei porscopini di 3 centimetri della Glazunkia Esterna. Ciascun porscopino ha sempre nove aculei (mentre i suoi cugini della Glazunkia Interna ne hanno sempre sette) e, pertanto, nove porscopini di due centimetri appuntati sul proprio corpo. Ciascuno di essi, a sua volta, ostenta nove porscopini da un centimetro da cui si dipartono nove porscopini da zero centimetri, ciascuno dei quali ha un naso. In totale si hanno $9 \times 9 \times 9 \times 1$ nasi, vale a dire che un porscopino di 3 centimetri della Glazunkia Esterna ha un potere d'acquisto di 729 nasi. Se ci fossimo trovati invece nella Glazunkia Interna e avessimo cominciato con un porscopino di quattro centimetri, esso avrebbe avuto un potere d'acquisto di $7 \times 7 \times 7 \times 7 \times 1$ ossia 2401 nasi.

Vediamo se possiamo trovare una regola generale per calcolare il potere d'acquisto (misurato in nasi) di ciascun porscopino. Si potrebbe avere:

Il potere d'acquisto di un porscopino con un dato numero di aculei e una determinata dimensione è:
se è di dimensione 0, sarà 1
altrimenti si trovi il potere d'acquisto di un porscopino con lo stesso numero di aculei ma di dimensione immediatamente inferiore a quella del porscopino dato e lo si moltiplichi per il numero degli aculei.

Possiamo abbreviare la regola servendoci di una notazione simbolica. Sia «a» il numero degli aculei c «d» la dimensione. Poniamo poi che «cond» stia per «se» e «t» per «altrimenti». Si usi infine una specie di notazione algebrica condensata in cui i nomi delle operazioni vengano messi a sinistra dei loro operandi all'interno di parentesi. Otteniamo allora:

```
(potere d'acquisto a d) è:
    cond (eq d 0) 1;
         t (times
             a
             (potere d'acquisto
                 a
                 (predecessore d)))
```

Questa è una esatta traduzione della precedente regola in una forma leggermente più simbolica. Possiamo renderla un po' più compatta e simbolica adottando un paio di nuove convenzioni. Poniamo ciascuno dei due casi (quello in cui d è uguale a 0 e il caso «altrimenti») tra parentesi; in generale, si usino le parentesi per racchiudere ciascuna unità logica. Si indichi poi con le parole «def» e «lambda» che si tratta della definizione di una nozione generale detta «potere d'acquisto», con due variabili (numero degli aculei a e dimensione d). Abbiamo allora la seguente definizione:

```
(def potere-di-acquisto
 (lambda (a d)
  (cond ((eq d 0) 1)
        (t
         (times a
(potere-di-acquisto a
            (predecessore d)))))))
```

Abbiamo detto prima che il potere d'acquisto di un porscopino a nove aculei di 3 centimetri è 729 nasi. La cosa si può esprimere nel modo seguente: (potere-di-acquisto 9 3) è uguale a 729. Analogamente (potere-di-acquisto 7 4) è uguale a 2401.

Torniamo al Lisp. Quasi alla fine dell'articolo del mese scorso vi avevo proposto un rompicapo che richiedeva di scrivere una funzione Lisp che riassumesse una famiglia di funzioni come «quadrato», «cubo», «4ª potenza», «5ª potenza» e così via. Vi chiedevo di riuscire a definire una funzione *generale* chiamata «potenza» con due variabili, tale che «(potenza 9 3)» dia 729, «(potenza 7 4)» dia 2401 e così via. Avevo presentato una «torre di potenze», cioè una torre infinitamente alta di definizioni Lisp distinte, una per ciascuna potenza, all'interno delle quali ogni volta c'è un riferimento alla potenza precedente. Allora un piano tipico di questa torre potrebbe essere:

```
(def 102-esima-potenza
 (lambda (a)
  (times a (101-esima-potenza a))))
```

Ovviamente la $101^{esima}$ potenza si riferirebbe alla $100^{esima}$ nella sua definizione e così via creando in tal modo un regresso piuttosto lungo fino ad arrivare al caso «embrionale» o più semplice. Sia detto per inciso: il caso più semplice non è «quadrato» o «prima potenza» ma

```
(def 0-esima-potenza (lambda (a) 1))
```

Ho detto che avevate tutte le informazioni necessarie per mettere insieme la definizione giusta. Dovevate soltanto preoccuparvi, ovviamente, di basare ogni piano della torre su quello che è il «predecessore immediato» o «predecessore» (tranne nel caso del piano terra che è un piano «che si sostiene da solo»). Per «predecessore» intendo:

```
(def predecessore
 (lambda (d) (difference d 1)))
```

Allora «(predecessore 102)» è 101. In realtà il Lisp ha un nome standard per questa funzione, precisamente «sub 1» e un nome anche per la sua inversa, la funzione successore, cioè «add 1». Se mettiamo insieme tutte queste osservazioni otteniamo la seguente definizione universale:

```
(def potenza
 (lambda (a d)
  (cond ((eq d 0) 1)
        (t
         (times a
         (potenza a
             (predecessore d)))))))
```

Questa è la risposta al rompicapo che vi avevo posto. Ehm, è buffo... Ho una stranissima impressione di *déjà vu*. Come mai?

La definizione che ho presentato è detta *ricorsiva* perché nella definizione si usa ciò che si deve definire, ossia si definisce qualcosa nei termini di se stesso. La cosa sarebbe considerata circolare, per non dire del tutto scorretta, da parte di chiunque. Per rendersi conto se il genietto del Lisp guarda con sospetto tale furberia, chiediamogli di eseguire (potenza 9 3):

```
—> (potenza 9 3)
729
—>
```

Incredibile! Niente da dire? Nessuno shock? Come può il genietto del Lisp avallare un tale nonsenso?

La migliore spiegazione che io possa dare è che la cosa non comporta, in realtà, nessuna circolarità. Anche se è vero che la definizione di «potenza» usa la parola «potenza», le due occorrenze si riferiscono a circostanze differenti. In poche parole, (potenza a d) è definito in termini di un caso più semplice, vale a dire (potenza a (predecessore d)). Sto allora definendo la $44^{esima}$ potenza in termini di $43^{esima}$ potenza e quest'ultima in termini della potenza immediatamente inferiore e così via discendendo la linea fino alla «base», come la chiamo io: la $0^{esima}$ potenza, che non ha bisogno di alcuna recursione. È sufficiente dire al geniettoche il suo valore è 1, cosa che abbiamo puntualmente fatto. Quindi, a guardar bene, questa definizione ricorsiva non è più circolare di quel che fosse la «torre di potenze» e non ne potete trovare un'altra più corretta. Questa definizione compatta è proprio quello che ci vuole per contenere l'intera «torre di potenze» in un'espressione finita. Ben lungi dall'essere circolare, è un maneggevole compendio di molte definizioni differenti, appartenenti a una medesima famiglia.

Se questo gioco di destrezza vi lascia ancora un po' scettici, provate a vedere nella figura in alto quello che farà il genietto Lisp se richiedete una «traccia» della funzione e poi chiedete ancora una volta di valutare (potenza 9 3).

Sulle righe segnate «ENTERING» il genietto Lisp stampa i valori dei due argomenti e sulle righe contrassegnate da «EXITING» stampa il valore che ha calcolato. Per ogni riga ENTERING c'è naturalmente una riga EXITING e le due sono allineate verticalmente, cioè sono scritte con una uguale rientranza rispetto al margine sinistro.

Come si vede, per trovare il valore di (potenza 9 3) il genietto deve prima calcolare (potenza 9 2). Quest'ultimo, però, non è un dato: bisogna conoscere il valore di (potenza 9 1) e questo, a sua volta, richiede la conoscenza di (potenza 9 0). Questo, finalmente, c'è stato dato: è 1. Adesso possiamo risalire, ricordando che per ottenere una risposta dalla risposta «più profonda» dobbiamo moltiplicare per 9. Abbiamo quindi 9, poi 81, poi 729.

Naturalmente non siamo noi ma il genietto che deve seguire questa via. Il genietto Lisp deve essere in grado di sospendere un calcolo per lavorare su un altro la cui risposta sia stata richiesta per prima. E anche il secondo calcolo può richiedere la risposta a un terzo, e così via ricorsivamente. A un certo punto, però, il gioco si ferma - cioè si arriva a un punto in cui un processo si completa e restituisce un valore - e questo mette in grado altri processi che si sono accumulati di restituire infine un valore; proprio come un gruppo di aereoplani che abbiano girato per ore e infine riescono ad atterrare, con ogni atterraggio che apre la strada a un altro atterraggio.

```
—> (potenza 9 3)
    ENTERING potenza (a = 9, d = 3)
      ENTERING potenza (a = 9, d = 2)
        ENTERING potenza (a = 9, d = 1)
          ENTERING potenza (a = 9, d = 0)
          EXITING    potenza (valore: 1)
        EXITING    potenza (valore: 9)
      EXITING    potenza (valore: 81)
    EXITING    potenza (valore: 729)
729
—>
```

***Il genio Lisp valuta (potenza 9 3)***

Di solito, il genietto Lisp non stampa una traccia di ciò che sta pensando finché non glielo si chiede. Che glielo si chieda o meno, comunque, questo è quello che succede dietro le quinte ogni volta che viene valutata una richiesta di funzione. Il saper affrontare queste definizioni ricorsive senza scomporsi è uno degli aspetti più positivi del Lisp.

Non sono così ingenuo da aspettarmi che ora abbiate capito tutto della ricorsività e possiate mettervi facilmente a scrivere enormi programmi ricorsivi. In effetti, la ricorsività può essere uno strumento straordinariamente sottile e a volte anche un esperto può trovarsi in difficoltà a cogliere il significato di una complicata definizione ricorsiva. Ho pensato, quindi, di farvi fare un po' di pratica con la ricorsività.

Facciamo un semplice esempio basato su questo indovinello: come si fa una pila di 13 pietre? Una risposta banale sarebbe: si mette una pietra su una pila di 12 pietre. Suponiamo di voler costruire una funzione Lisp che ci dia non una pila di 13 pietre ma una lista formata da 13 copie dell'atomo «pietra» o, in generale, da *n* copie di quell'atomo. Possiamo fondare la nostra risposta sulla risposta ricorsiva, corretta anche se apparentemente banale, all'indovinello. La nozione generale è quella di costruire la risposta per *n* a partire dalla risposta per il predecessore di *n*. Costruire come? Usando la funzione cons che serve a costruire liste, ecco come. Qual è il caso embrionale? Cioè, per quale valore di *n* l'indovinello non presenta assolutamente alcun problema? È facile: quando *n* è uguale a 0, la nostra lista dovrebbe essere vuota, il che significa che la risposta è nil. Possiamo ora mettere insieme le nostre osservazioni in questo modo:

```
(def mucchio-di-pietre
 (lambda (n)
  (cond ((eq n 0) nil)
        (t
         (cons 'pietra
               (mucchio-di-pietre
                (predecessore n)))))))
```

Ora osserviamo il genietto mentre mette insieme un mucchio di pietre molto piccolo (con «traccia» inserita, tanto per divertirsi). Basta osservare la figura della pagina seguente per vedere quello che succede.

Abbiamo così una chiara illustrazione di ciò che si chiama «costruire una lista usando cons». Ora facciamo un'altra prova, questa volta con un vecchio cavallo di battaglia del Lisp e della ricorsività in genere. Guardate la definizione e vedete se riuscite a immaginare a che cosa dovrebbe servire; poi continuate a leggere per controllare se avete visto giusto.

```
(def wow
 (lambda (n)
  (cond ((eq n 0) 1)
        (t
         (times n (wow (sub 1 n)))))))
```

Ricordate che «sub 1» ha lo stesso significato di «predecessore». Così per scherzo, perché non calcolate il valore di (wow 100)? (Se stamattina avete fatto una bella colazione, potete provare a fare il calcolo a mente.)

Succede spesso che i genietti Lisp meditino ad alta voce mentre eseguono i desideri e mi è appena accaduto di sentirne uno mentre eseguiva il desiderio «(wow 100)». Il suo ragionamento era pressappoco così:

«Hmm... (wow 100), eh? Bene, 100 non è sicuramente uguale a 0, quindi penso che la risposta debba essere 100 volte ciò che sarebbe stata se il problema fosse stato (wow 99). Oh, mi sa che è una cosa da nulla. Vediamo, 99 è uguale a 0? No, non si direbbe, quindi penso che la risposta a *questo* problema debba essere 99 volte ciò che sarebbe stata se il problema fosse stato (wow 98). Vediamo...»

A questo punto l'autore, avendo degli affari urgenti da sbrigare in banca, aveva abbandonato l'allegro genietto e non era ritornato che qualche millisecondo più tardi. Il genietto stava proprio finendo e diceva:

«... e ora non mi rimane che moltiplicare *quello* per 100 e avrò il mio risultato finale. Facile come bere un bicchiere d'acqua! Direi che la risposta sia 93326 -

```
—> (mucchio di pietre 2)
    ENTERING mucchio di pietre (n = 2)
      ENTERING mucchio di pietre (n = 1)
        ENTERING mucchio di pietre (n = 0)
        EXITING   mucchio di pietre (valore: nil)
      EXITING   mucchio di pietre (valore: (pietra))
    EXITING   mucchio di pietre (valore: (pietra pietra))
(pietra pietra)
—>
```

***Il genio mette insieme un mucchietto di pietre***

2154439441526816992388562667004-907159682643816214685929638952 1-759999322991560894146397615651 8-286253697920827223758251185210 9-16864000000000000000000000000, se non mi sbaglio.»

È la risposta che avevi trovato, caro lettore? Ah, vedo dove hai sbagliato: è quando hai moltiplicato per 52. Torna indietro e riprova da quel punto; e cerca di fare più attenzione quando sommi quelle lunghe colonne. Sono sicuro che questa volta troverai la risposta giusta.

Questa funzione «wow» è chiamata di solito «fattoriale»; n fattoriale è abitualmente definito come il prodotto di tutti i numeri da 1 a n. Una definizione ricorsiva, invece, guarda le cose in modo leggermente differente; parlando ricorsivamente, n fattoriale è semplicemente il prodotto di n e del fattoriale precedente. Questo riduce il problema a uno più semplice dello stesso tipo. Quello più semplice sarà a sua volta ridotto, e così giù giù, fino ad arrivare al problema più semplice di quel tipo, che io chiamo caso embrionale o base.

Un fumetto apparso qualche anno fa sul «New Yorker» illustra perfettamente il concetto. In esso si vede un uomo di circa 50 anni che tiene in mano una fotografia di se stesso presa circa 10 anni prima. In quella fotografia tiene in mano una fotografia di se stesso presa 10 anni prima di *quella*. E così via, finché alla fine si tocca il fondo (quasi alla lettera) in una fotografia di un robusto neonato col sedere in aria. Quest'idea delle fotografie ricorsive che documentano la vostra crescita è molto attraente e vorrei che i miei genitori vi avessero pensato. Fate un raffronto con il più famoso regresso all'infinito sul pacchetto del Morton Salt, in cui una ragazza tiene in mano una scatola di Morton Salt su cui è raffigurata la sua immagine. Poiché la ragazza dell'immagine non è più giovane dell'altra, non c'è nessun punto d'arrivo e il regresso è, almeno teoricamente, all'infinito.

Un'impostazione di tipo ricorsivo funziona quando si ha una famiglia di problemi tra loro connessi, almeno uno dei quali può avere una soluzione immediata. Questo è quello che io chiamo il caso embrionale. (Nell'esempio del fattoriale è il caso «(eq n 0)», la cui soluzione è 1.) Ogni problema può venir considerato come caso particolare di un problema generale («Come si calcola un fattoriale?»). La ricorsività si basa sul fatto che le soluzioni a vari casi sono collegate le une alle altre in qualche forma logica. (Per esempio, io potrei dirvi molto agevolmente il valore di 100 fattoriale se solo qualcuno mi passasse il valore di 99 fattoriale; tutto quello che ho bisogno di fare è moltiplicare per 100.) Si potrebbe dire che il motto di chi ricorre a procedimenti ricorsivi sia: «Per Bacco! Vi risolverei *questo* caso se soltanto qualcuno mi facesse magicamente avere la soluzione del caso che è di un passo più vicino al caso embrionale». Ovviamente questo motto presuppone che alcuni casi siano, in qualche senso, «più vicini» al caso embrionale di altri. Presuppone, cioè, che ci sia un cammino naturale che porta da un caso qualsiasi giù fino al caso embrionale, un cammino i cui passi siano chiaramente distinguibili.

In tutte le circostanze questo è il presupposto più ragionevole. Per far emergere la vera natura di questo cammino guida della ricorsività dovete rispondere a due Domande Fondamentali:

1. Qual è il caso embrionale?
2. Che relazione c'è tra un caso tipico e il caso immediatamente più semplice?

In realtà ciascuna di queste Domande Fondamentali si divide in due sottodomande (come conviene a ogni domanda ricorsiva che si rispetti) una riguardante il come si riconosce dove siete e come ci si deve muovere e l'altra riguardante la soluzione a ogni stadio dato. Allora, messe in termini più espliciti, le nostre Domande Fondamentali sono:

1*a*. Come si fa a sapere quando si è raggiunto il caso embrionale?
1*b*. Qual è la risposta embrionale?

2*a*. Partendo da un caso tipico, come si fa a compiere esattamente un passo verso il caso embrionale?
2*b*. Come si costruisce la risposta a questo caso partendo dalla risposta «magicamente data» al caso più semplice?

La domanda 2*a* riguarda la natura della *discesa* verso il caso embrionale, la base. La domanda 2*b* riguarda l'aspetto inverso, vale a dire l'*ascesa* che vi riporta dal basso ai livelli superiori.

Nel caso del fattoriale, le risposte alle Domande Fondamentali sono:

1*a*. Il caso embrionale si ha quando l'argomento è 0.
1*b*. La risposta embrionale è 1.

2*a*. Sottrarre 1 dall'attuale argomento.
2*b*. Moltiplicare la risposta «magica» per l'attuale argomento.

Si noti come le risposte a queste quattro domande siano tutte incorporate nella definizione ricorsiva di «wow».

La ricorsività si fonda sul presupposto che presto o tardi si arrivi alla base. Un modo per essere *sicuri* di arrivare alla base è quello che tutti i passi «discendenti» muovano nella stessa direzione allo stesso ritmo, in modo che il percorso sia ovviamente lineare. Per esempio, è ovvio che continuando a sottrarre 1 raggiungerete alla fine 0, sempre che siate partiti con un intero positivo. È altrettanto ovvio che applicando l'operazione «cdr» per l'abbreviazione di liste raggiungerete alla fine nil, sempre che siate partiti con una lista finita. Per questa ragione, sono comuni i procedimenti ricorsivi che utilizzano «sub 1» o «cdr» per definire il loro percorso di discesa verso il basso. Presenterò più avanti un percorso ricorsivo basato sull'operazione cdr, ma prima voglio mostrare un divertente processo ricorsivo numerico in cui il percorso verso il caso embrionale è tutt'altro che lineare.

Consideriamo il famoso problema «3n+1»: si parte con un qualsiasi intero positivo e, se è pari, lo si dimezza, altrimenti lo si moltiplica per 3 e si aggiunge 1. Chiamiamo il risultato di questa operazione su n «(dotpu n)» (che sta per «dimezzare o triplicare più uno»). Ecco una definizione Lisp di dotpu:

```
(def dotpu
 (lambda (n)
  (cond ((pari n) (dimezzare n))
        (t (add 1 (times 3 n))))))
```

Questa definizione presuppone che altrove siano state definite o saranno definite altre due funzioni, cioè «pari» e «dimezzare» («add 1» e «times» sono, come ho già detto, parti intrinseche del Lisp). Ecco le definizioni mancanti:

```
(def pari
 (lambda (n) (eq (remainder n 2) 0)))

(def dimezzare (lambda (n) (quotient n
2)))
```

Cosa pensate che avvenga se iniziate con un intero e continuate ad applicare dotpu? Prendiamo 7, per esempio, come punto di partenza. Prima di calcolare, provate a indovinare che tipo di comportamento potrebbe risultare.

Risulta che il percorso è spesso sorprendentemente caotico e accidentato. Per esempio, se iniziamo con 7, il procedimento ci porta a 22, poi 11, poi 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1, 4, 2, 1, 4, 2, 1... Si noti che finiamo in uno stretto loop. Supponiamo, allora, di metterci d'accordo sul fatto che se raggiungiamo 1, abbiamo raggiunto il fondo e possiamo fermarci. Si potrebbe chiedere: «Chi ci dice che raggiungeremo 1? È garantito?» Effettivamente, se non avete mai provato, non c'è alcuna ovvia ragione per credere che si raggiungerà 1. (Nel caso di 7, vi aspettavate che succedesse?) La sperimentazione numerica, però, rivela che il processo ha una notevole affidabilità; parrebbe che, da qualsiasi punto si parta, dopo un po' si raggiunga sempre 1. (Provate a partire con 27 se volete un vero percorso da montagne russe!)

Sapreste scrivere una funzione ricorsiva per mostrare il percorso seguito da un punto di partenza arbitrario «giù» fino a 1? Si noti che dico «giù» a ragion veduta, visto che molti dei passi sono in realtà *su*. Così il percorso che parte da 3 sarebbe la lista «(3 10 5 16 8 4 2 1)». Per risolvere questo rompicapo bisogna che torniate indietro e rispondiate da soli alle due Domande Fondamentali della Ricorsività, nella loro applicazione a questo caso. Notate:

```
(cond ((not (volete aiuto))
       (not (leggete oltre)))
      (t (leggete oltre ))).
```

Prima di tutto, il caso embrionale. Questo è facile. È già stato definito come l'arrivo a 1 e la risposta embrionale, la più semplice possibile, è la lista «(1)», un percorso breve ma valido che inizia e termina con 1.

Poi, alcuni casi più tipici. Che operazione ci porterà da 7 più vicino di un passo all'embrionale 1? Certamente non l'operazione «sub 1». No, per definizione è la funzione dotpu stessa che vi porta sempre «più vicino» a 1, anche quando vi porta *su*. La singolarità dell'esempio sta naturalmente in questa sconcertante caratteristica. E a proposito di 2*b*, come costruire ricorsivamente una lista che documenti il nostro percorso ad alta oscillazione? Bene, il percorso di 7 si ottiene considerando il percorso più breve di (dotpu 7), ossia 22. Dopo tutto, 22 è un passo più vicino di 7 all'embrione!

Queste risposte ci mettono in grado di scrivere la richiesta definizione di funzione, usando «oeso» come variabile prestanome, dove «oeso» è un ben noto acronimo per «oeso (e solo oeso)», che si espande ricorsivamente a «oeso (e solo oeso) (e solo oeso (e solo oeso))», e così via.

```
(def percorso-a-1
 (lambda (oeso)
  (cond ((eq oeso 1) '(1))
        (t
         (cons oeso
               (percorso-a-1
                (dotpu oeso)))))))
```

Nella figura qui sotto si può vedere il modo in cui il genietto Lisp «pensa» (evidenziato inserendo l'istruzione «traccia»).

Si noti l'assoluta regolarità (forma a *V*) del margine sinistro del tracciato, nonostante la caoticità dei numeri in gioco. Non tutti i processi ricorsivi mostrano questa bellezza geometrica quando vengono tracciati. La ragione è che alcuni problemi richiedono la soluzione di *più di un sotto-problema*. Per considerare un esempio pratico, della vita quotidiana, di problema di questo genere, consideriamo il modo in cui potreste contare tutti gli unicorni d'Europa. Non è certo un'impresa da poco, eppure c'è un'elegante risposta ricorsiva. Contiamo tutti gli unicorni del Portogallo (il «car» d'Europa, per così dire), poi contiamo tutti gli unicorni delle altre nazioni d'Europa (il «cdr» d'Europa, per continuare con la metafora) e infine sommiamo i due risultati.

```
—> (percorso-a-1 3)
ENTERING   percorso-a-1    (oeso = 3)
  ENTERING   percorso-a-1    (oeso = 10)
    ENTERING   percorso-a-1    (oeso = 5)
      ENTERING   percorso-a-1    (oeso = 16)
        ENTERING   percorso-a-1    (oeso = 8)
          ENTERING   percorso-a-1    (oeso = 4)
            ENTERING   percorso-a-1    (oeso = 2)
              ENTERING   percorso a-1-    (oeso = 1)
              EXITING    percorso-a-1    (valore:   (1))
            EXITING    percorso-a-1    (valore:   (2 1))
          EXITING    percorso-a-1    (valore:   (4 2 1))
        EXITING    percorso-a-1    (valore:   (8 4 2 1))
      EXITING    percorso-a-1    (valore:   (16 8 4 2 1))
    EXITING    percorso-a-1    (valore:   (5 16 8 4 2 1))
  EXITING    percorso-a-1    (valore:   (10 5 16 8 4 2 1))
EXITING    percorso-a-1    (valore:   (3 10 5 16 8 4 2 1))
(3 10 5 16 8 4 2 1)
—>
```

***Come «pensa» il genio quando esegue un tracciato***


**TEMI METAMAGICI**

**LE SCIENZE** *edizione italiana di* **SCIENTIFIC AMERICAN**

*tra gli articoli di Douglas Hofstadter, il titolare della rubrica mensile «Temi metamagici» vi segnaliamo:*

**CON QUESTO TITOLO SI PRESENTA UN NUOVO AUTORE DELLA RUBRICA** (n. 152)

**INGANNI DEL PRINCIPIO DI INDETERMINAZIONE E PARADOSSI DELLA MECCANICA QUANTISTICA** (n. 157)

**COME POSSONO I CALCOLATORI COMPRENDERE L'ANALOGIA, IL NUCLEO DEL PENSIERO UMANO?** (n. 159)

**«NATIONAL ENQUIRER» E «THE SKEPTICAL INQUIRER»** (n. 165)

**IL CODICE GENETICO È ARBITRARIO, OSSIA POTREBBE FUNZIONARE COSÌ BENE UN ALTRO CODICE?** (n. 166)

**INSENSIBILITÀ NUMERICA: PERCHÉ L'IGNORANZA NUMERICA PUÒ ESSERE ALTRETTANTO PERICOLOSA DELL'IGNORANZA LINGUISTICA** (n. 168)

**IL NOMIC: UN GIOCO ESALTANTE CHE ESPLORA LA RIFLESSIVITÀ DELLA LEGGE** (n. 169)

**TAGLIO, SFOGGIO, HRUSKA, EVOLUZIONE DEL COMPORTAMENTO E ALTRI GIOCHI DI STRATEGIA** (n. 172)

**VARIAZIONI SU UN TEMA COME ESSENZA DELL'IMMAGINAZIONE** (n. 173)

**«PRESUPPOSTI RIDUTTIVI» E LORO EFFETTI SULLA SCRITTURA E SUL PENSIERO** (n. 174)

**ENUNCIATI SIMILI A VIRUS E STRUTTURE AUTOREPLICANTESI** (n. 175)

Si noti come questo produca due più piccoli sottoproblemi, che a loro volta produrranno ciascuno due sottoproblemi, e così via. Per esempio, come si fa a contare tutti gli unicorni del Portogallo? È facile: basta sommare il numero di unicorni dell'Estremadura (il «car» del Portogallo) con il numero di unicorni del resto del Portogallo (il «cdr» del Portogallo). E come si fa a contare gli unicorni dell'Estremadura (per non parlare di quelli delle altre regioni del Portogallo)? Con ulteriori divisioni, naturalmente. Qual è, allora, la base? Bene, le regioni possono essere divise in distretti, i distretti in chilometri quadrati, i chilometri quadrati in ettari, gli ettari in metri quadrati, e qui possiamo anche fermarci senza divisioni ulteriori.

Anche se la cosa può sembrare piuttosto ardua, per compiere un censimento totale non c'è davvero altro modo che attraversare ogni singola parte a ogni livello dell'intera struttura, per quanto grande essa sia. C'è una perfetta corrispondenza tra questo censimento di unicorni e il problema Lisp di stabilire quanti atomi ci siano in una lista arbitraria. Come possiamo scrivere una funzione Lisp chiamata «atomcount» che ci dia come risposta 15 per la strana lista seguente (che chiameremo «brahma»)?

```
(((ac ab cb) (ba bc ac))
          ab
 ((cb ca ba) cb (ac ab cb)))
```

Un metodo, espresso ricorsivamente, è esattamente analogo a quello per accertare la popolazione di unicorni in Europa. Vedete un po' se riuscite a formularlo.

L'idea è questa. Vogliamo costruire la risposta - cioè 15 - partendo dalle risposte a più semplici problemi di calcolo di atomi. Bene, è ovvio che un problema di calcolo di atomi più semplice di (atomcount brahma) è (atomcount (car brahma)). Un altro è (atomcount(cdr brahma)). Le risposte a questi due problemi sono rispettivamente 7 e 8. Ora, 15 è chiaramente costituito dalla somma di 7 e 8 - il che è sensato, dopo tutto, visto che il numero totale di atomi deve essere il numero del car più il numero del cdr; non c'è posto in cui possano nascondersi altri atomi. Questa analisi ci dà la seguente definizione ricorsiva, con «s» come variabile prestanome:

```
(def atomcount
 (lambda (s)
  (plus (atomcount (car s))
        (atomcount (cdr s)))))
```

Sembra semplice, ma ci sono un paio di difetti. Prima di tutto, abbiamo scritto la parte *ricorsiva* della definizione, ma abbiamo totalmente dimenticato l'altra metà, altrettanto vitale, la base. Mi ricorda quel giudice del Maryland che, come riportava il giornale, aveva decretato che «un cavallo è un animale a quattro zampe generato da altri due cavalli». È un'ottima definizione, ma su cosa si fonda? Lo stesso si può dire di atomcount: qual è il caso più semplice, il caso embrionale, di atomcount? È quando ci viene richiesto di contare gli atomi di un singolo atomo. La risposta in questo caso è naturalmente 1. Ma come facciamo a sapere di aver di fronte un atomo? Per fortuna, il Lisp ha in sé una funzione detta «atom» che ha valore t (per «true», cioè «vero») ogni volta che si ha un atomo e in caso contrario nil. Così «(atom 'plop')» ci dà t e «(atom '(a b c))» ci dà nil. Possiamo allora risistemare la nostra definizione:

```
(def atomcount
 (lambda (s)
  (cond ((atom s) 1)
        (t
         (plus (atomcount (car s))
               (atomcount (cdr s)))))))
```

Ma non va ancora bene del tutto. Se richiediamo al genietto l'atomcount di «(a b c)», invece di 3 otterremo come risposta 4. Incredibile! Cosa avviene? Possiamo affrontare il problema cercando un esempio ancora più semplice: se chiediamo (atomcount '(a)), otteniamo 2 invece di 1. Ora dovrebbe essere più chiaro qual è l'errore: 2 = 1+1, dove 1 è fornito sia dal car sia dal cdr di «(a)». Il car è l'atomo «a», che dovrebbe essere effettivamente calcolato 1, ma il cdr è nil, che non dovrebbe essere calcolato. Come mai, allora, nil dà un atomcount di 1? Perché nil non è soltanto una lista vuota ma anche un atomo! Per sopprimere questo effetto negativo inseriamo semplicemente un'altra clausola cond in alto:

```
(def atomcount
 (lambda (s)
  (cond ((null s) 0)
        ((atom s) 1)
        (t
         (plus (atomcount (car s))
               (atomcount (cdr s)))))))
```

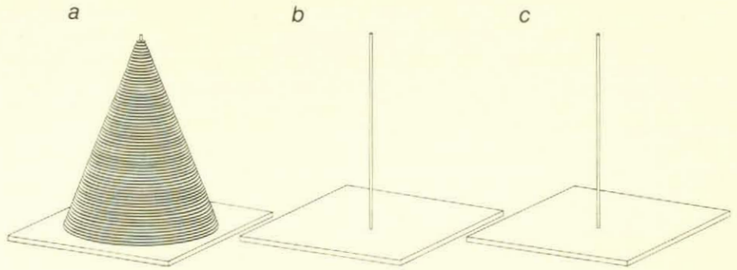
«(null s)» è semplicemente un altro modo per dire «(eq s nil)». In generale, quando si vuole stabilire se il valore di un'espressione è o non è nil, si può usare la funzione «null», che dà t se sì e nil se no. Quindi, per esempio, «(null (null nil)» dà come valore nil, dato che la funzione interna dà come valore t, e t non è nil! E «(null '(null nil))»?

Ora, cosa avviene quando applichiamo la nostra funzione a brahma, il suo obiettivo originale? Nella pagina a fronte si può vedere il risultato che si ottiene.

```
—> (atomcount brahma)
   ENTERING atomcount
      (s = (((ac ab cb) ac (ba bc ac)) ab ((cb ca ba) cb (ac ab cb))))
     ENTERING atomcount (s = ((ac ab cb) ac (ba bc ac)))
       ENTERING atomcount (s = (ac ab cb))
         ENTERING atomcount (s = ac)
         EXITING    atomcount (valore: 1)
         ENTERING atomcount (s = (ab cb))
           ENTERING atomcount (s = ab)
           EXITING    atomcount (valore: 1)
           ENTERING atomcount (s = (cb))
             ENTERING atomcount (s = cb)
             EXITING    atomcount (valore: 1)
             ENTERING atomcount (s = nil)
             EXITING    atomcount (valore: 0)
           EXITING atomcount (valore: 1)
         EXITING atomcount (valore: 2)
       EXITING    atomcount (valore: 3)
       ENTERING atomcount (s = (ac bc ac)))
         ENTERING atomcount (s = ac)
         EXITING    atomcount (valore: 1)
         ENTERING atomcount (s = ((ba bc ac)))
           ENTERING atomcount (s = (ba bc ac))
             ENTERING atomcount (s = ba)
             EXITING    atomcount (valore: 1)
             ENTERING atomcount (s = (bc ac))
               ENTERING atomcount (s = bc)
               EXITING    atomcount (valore: 1)
               ENTERING atomcount (s = (ac))
                 ENTERING atomcount (s = ac)
                 EXITING    atomcount (valore: 1)
                 ENTERING atomcount (s = nil)
                 EXITING    atomcount (valore: 0)
               EXITIng atomcount (valore: 1)
             EXITING atomcount (valore: 2)
           EXITING    atomcount (valore: 3)
           ENTERING atomcount (s = nil)
           EXITING    atomcount (valore: 0)
         EXITING atomcount (valore: 3)
       EXITING atomcount (valore: 4)
     EXITING    atomcount (valore: 7)
     ENTERING atomcount
        (s = (ab (cb ca ba) cb (ac ab cb))))
       ENTERING atomcount (s = ab)
       EXITING    atomcount (valore: 1)
       ENTERING atomcount
          (s = (((cb ca ba) cb (ac ab cb))))
         ENTERING atomcount (s = ((cb ca ba) cb (ac ab cb)))
           ENTERING atomcount (s = cb ca ba))
             ENTERING atomcount (s = cb)
             EXITING    atomcount (valore: 1)
             ENTERING atomcount (s = (ca ba))
               ENTERING atomcount (s = ca)
               EXITING    atomcount (valore: 1)
               ENTERING atomcount (s = (ba))
                 ENTERING atomcount (s = ba)
                 EXITING    atomcount (valore: 1)
                 ENTERING atomcount (s = nil)
                 EXITING    atomcount (valore: 0)
               EXITING atomcount (valore: 1)
             EXITING atomcount (valore: 2)
           EXITING    atomcount (valore: 3)
           ENTERING atomcount (s = (cb (ac ab ch)))
             ENTERING atomcount (s = cb)
             EXITING    atomcount (valore: 1)
             ENTERING atomcount (s = ((ac ab cb)))
               ENTERING atomcount (s = (ac ab cb))
                 ENTERING atomcount (s = ac)
                 EXITING    atomcount (valore: 1)
                 ENTERING atomcount (s = (ab cb))
                   ENTERING atomcount (s = ab)
                   EXITING    atomcount (valore: 1)
                   ENTERING atomcount (s = (cb))
                     ENTERING atomcount (s = cb)
                     EXITING    atomcount (valore: 1)
                     ENTERING atomcount (s = nil)
                     EXITING    atomcount (valore: 0)
                   EXITING atomcount (valore: 1)
                 EXITING atomcount (valore: 2)
               EXITING    atomcount (valore: 3)
               ENTERING atomcount (s = nil)
               EXITING    atomcount (valore: 0)
             EXITING atomcount (valore: 3)
           EXITING atomcount (valore: 4)
         EXITING    atomcount (valore: 7)
         ENTERING atomcount (s = nil)
         EXITING    atomcount (valore: 0)
       EXITING atomcount (valore: 7)
     EXITING atomcount (valore: 8)
   EXITING atomcount (valore: 15)
15
—>
```

*La funzione Lisp «atomcount» applicata alla lista «brahma»*






*Il rompicapo della Torre di Brahma. I 64 dischi su «a» devono essere collocati nello stesso ordine su «b»*

Si noti la più complessa topografia di questo processo ricorsivo, con i suoi alterni spostamenti dal margine sinistro. Il precedente processo ricorsivo a forma di *V* sembrava una semplice discesa in un canyon con le pareti smussate e la successiva risalita; quest'ultimo, invece, ci fa pensare alla discesa e risalita di un canyon più tormentato, in cui incontriamo vari «sotto-canyon» che dobbiamo percorrere allo stesso modo, e chi sa quanti livelli di tale struttura saremo chiamati ad affrontare nella nostra esplorazione? La forma descritta da una struttura che procede indefinitamente come quella è stata chiamata «*fractal*» da Benoit Mandelbrot.

Si noti che in questo processo ricorsivo abbiamo più di un caso embrionale («null» e «atom») e più di un modo per scendere verso il caso embrionale (via car o via cdr). Le nostre Domande Fondamentali possono essere allora un po' riviste:

1*a*. C'è un solo caso embrionale o ce ne sono parecchi o ce n'è addirittura una classe infinita?

1*b*. Come si fa a sapere quando si è raggiunto un caso embrionale?

1*c*. Quali sono le risposte ai vari casi embrionali?

2*a*. Da un caso tipico c'è esattamente un modo per muovere verso un caso embrionale o ci sono varie possibilità?

2*b*. Da un caso tipico, come stabilire quale prendere delle varie strade verso un caso embrionale?

2*c*. Come costruire la risposta a questo caso partendo dalle risposte «magicamente date» a uno o più casi più semplici?

Uno dei più eleganti processi ricorsivi che conosca trae origine dal famoso rompicapo noto con i nomi di «Torre di Lucas», «Torre di Hanoi» e «Torre di Brahma». Inventato dal matematico francese Edouard Lucas nel XIX secolo, è legato a questa leggenda popolare:

Nel grande tempio di Brahma a Benares, su un piatto d'ottone posto sotto la cupola che segna il centro del mondo, ci sono 64 dischi di oro puro che i sacerdoti spostano, uno alla volta, da uno all'altro di tre perni di diamante secondo l'immutabile legge di Brahma: nessun disco può essere posto su un disco più piccolo. All'inizio del mondo tutti e 64 i dischi formavano la Torre di Brahma su un perno. Ora, invece, è in svolgimento il processo di trasferimento della torre da un perno a un altro. Quando l'ultimo disco sarà infine a posto, formando di nuovo la Torre di Brahma ma su un perno differente, allora arriverà la fine del mondo e tutto tornerà in polvere.

Qui sotto si può vedere una raffigurazione del rompicapo; ho indicato con «a», «b» e «c» i tre perni.

Se ci pensate un po', scoprirete certamente il metodo sistematico che i sacerdoti devono seguire per spostare i dischi dal perno «a» al perno «b». Per soli tre dischi, per esempio, non è difficile scrivere l'ordine delle mosse:

ab ac bc ab ca cb ab

In questo caso, l'atomo Lisp «ab» rappresenta un passaggio dal perno «a» al perno «b». La struttura di ciò che avviene, però, non è rivelata da una semplice elencazione di tali atomi. Appare meglio se si raggruppano gli atomi nel modo seguente:

ab ac bc  ab  ca cb ab

Il primo dei tre gruppi serve a trasferire una torre da 2 dal perno «a» al perno «c», liberando un disco più grosso che, con la mossa centrale, viene spostato dal perno «a» al perno «b». Il terzo gruppo è del tutto simile al primo, in quanto riporta la torre da 2 dal perno «c» al perno «b». Riuscire a muovere tre dischi, quindi, dipende dalla capacità di muoverne due. Analogamente, riuscire a muovere 64 dischi dipende dalla capacità di muoverne 63. Vi basta? Ora cercate di scrivere una funzione Lisp che vi dia una soluzione alla torre di Brahma per *n* dischi. (Se preferite, potete marcare i tre perni con cifre invece che con lettere, in modo che le mosse siano rappresentate da numeri di due cifre come 12.) Darò la soluzione il mese prossimo, sempre che, naturalmente, i sacerdoti non riescano nel frattempo, lavorando giorno e notte, a raggiungere l'obiettivo di provocare la fine del mondo.